VOLUME XIII 1922

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

Published Monthly Except June to August by
WARWICK and York, Inc.,
York, Pa. *Baltimore, Md.*

TABLE OF CONTENTS

INDEX TO AUTHORS

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Vol. XIII | January, 1922 | No. 1

RESEARCH PROBLEMS RAISED IN RECENT ISSUES OF EDUCATIONAL PERIODICALS

LAURA ZIRBES

The Lincoln School of Teacher's College

There are numerous indices of the present trend in educational research. While no *one* index is sufficiently representative, a number of careful searches in well chosen but restricted areas should reveal not only the trend of research, but a series of *unsolved* problems. The solution of such problems is often a significant contribution to scientific progress and educational practice. They are most often formulated as a result of and in connection with studies published in periodicals, books and monographs, in the pursuit of research and inquiry in graduate schools and bureaus of research, in experimental schools and in oral utterances of educational leaders. They seldom stand out conspicuously, but are more often mentioned incidentally in connection with the presentation of a practical situation which calls for their solution. This article is one of a series which will, from time to time, seek to rescue such statements of problems from oblivion, in the hope that re-statement, frequency of occurrence and accessibility will combine to define the trend and, by speeding the attack, hasten their solution.

The search in this instance was confined to the three most recent issues of the following periodicals, except as noted:[1]

1. Educational Administration and Supervision
2. The Elementary School Journal
3. The Journal of Applied Psychology
4. The Journal of Educational Psychology

[1] The search was confined to autumn issues of 1921 because several publications have no summer numbers and the inclusion of materials appearing on widely separated and non-contiguous dates, and consequently not within the limits of single volumes, would be less significant for the purpose of defining a trend. Several carefully limited searches will serve that purpose better than one more inclusive investigation.

5. The Journal of Educational Research
6. The Psychological Bulletin
7. The School Review
8. School and Society (10 issues)
9. The Teachers College Record (1 issue)

Periodicals will be referred to by the numbers assigned in this list and page numbers refer to the current volume (1921). These are given so that those who are in a position to conduct research may readily find the original statement of any problem in its setting. Differences in form of statement are due to an attempt to retain the original phrasing whenever possible.

I. Problems Referring to Mental Tests

Studies of the constancy of the IQ. 6, p. 339, 341, 342; 4, p. 314, 323.

To what extent is mental age valuable for prognosis of pupil achievement in the lower grades? 4, p. 383.

There is need for the validation of Stanford norms, the revision of tests and administrative procedure on the basis of a large number of unselected children, by a group of disinterested psychologists. 4, p. 400.

Why is there a progressive increase in overlapping evident in individual test (Binet) results and absent in the case of group tests? 4, p. 405.

Comparison, analysis and evaluation of group tests of intelligence. 6, p. 342.

Need for determining the reliability and validity of tests and methods used in the interpretation of results. 5, p. 136.

Adequate determination of the total distribution of abilities in adolescence because school groups show effect of cumulative selection. 3, p. 70.

A number of sets of norms obtained from representative groups of cases. 3, p. 70.

Critical investigation of the concept "general intelligence" and investigation regarding the specificity of prognosis problems. 3, p. 76.

Investigation and qualitative examination of test responses in individual cases where results of retest show great inconstancy of IQ. 3, p. 158.

Investigation of the relation of mild social and physical maladjustment of superior individuals, to marked intellectual achievement. 8, p. 425.

Intensive scientific study of sixty or more carefully selected infants under controlled conditions and through a number of years to determine educability due to heredity; racial differences; health controls. 8, p. 312.

II. Curriculum Studies

Scientific determination of curricula in history and allied subjects. 5, p. 294; 7, p. 617; 7, p. 573; 8, p. 386.

Reconstruction of college curricula in terms of distribution of abilities and individual differences. 8, p. 389; 8, p. 437.

Experimental study to show how the double demand for assured values of racial experience and the proper utilization of children's purposes may best be met. 9, p. 287; 9, p. 289.

Curriculum research in chemistry. 7, p. 646; 8, p. 220.

Statement of objectives in reading and experimental determination of materials to accord with objectives. 7, p. 573.

Mathematical curricula for high schools based on investigation of socially valuable relations between quantities, and experimental data concerning training in the ability to think in terms of quantitative data of proven social significance. 7, p. 646.

Curriculum reconstruction in geography. 8, p. 437.

Vocabulary studies to ascertain the relative importance and significance of words of foreign derivation. 9, p. 368.

How may health instruction be co-ordinated with other subjects of the curriculum. 2, p. 41.

III. Studies Pertaining to Administration and Supervision

Construction of scientific instruments and methods that will aid in diagnosis of teaching and supervisory process and the selection and improvement of teachers. 5, p. 83; 3, p. 39.

Experimental investigation of effect of textbooks on outcomes of instruction. 4, p. 342; 9, p. 358.

Necessity for studying the effect of various marking systems and of selecting a system with reference to objectives and function in controlling instruction. 7, p. 510.

What is the maximum age and ability range of an effective group? 4, p. 342.

An application of statistical method in the analysis of factors of teaching success. 5, p. 89.

Suggestions as to necessary refinement of survey methods and interpretations. 1, 433.

Experimental evaluation of classification schemes: horizontal and graded versus vertical and parallel. 2, p. 71.

Experimental investigation of improvement in the quality of instruction due to controlled causes. 8, p. 469.

Occupational descriptions of university positions as one possible means for the improvement of instruction in universities. 8, p. 293.

IV. Educational Test Problems

Reformulation and analysis of silent reading problems as next steps. 4, p. 304; 4, p. 350; 6, p. 350; 8, p. 211.

Does the function represented by the Thorndike-McCall Reading Test yield to practice or training or is it one which develops primarily as a result of inner growth? 4, p. 384.

Research leading to the improvement of reading tests. 4, p. 464.

Need for additional forms of Gray Tests. 4, p. 381.

Necessity for determining reliability and validity of subject tests in order to interpret and use results wisely. 5, p. 136.

Vocabulary checks on material in standard tests by use of Thorndike's Wordbook. 9, p. 368.

V. Learning Studies

What are the methods employed by children in the gradual acquirement of the power of reading numerals? 4, p. 365.

Review of scattered data and experimental and statistical data to reveal to what extent various types of specific training with words increase vocabulary. 4, p. 456.

To what extent would an experimental evaluation of "piece-meal learning" affect curriculum research? 4, p. 474.

Experimental determination of the educational significance of individual differences upon which differentiation in curricula can be based. 5, p. 151.

Need for development of tests of pupils' use of economical and desirable methods of study, not to show results of study but *habits of study*. 7, p. 706.

Case studies to determine whether inefficient work is due to specific present disability or intelligence limitations. 5, p. 292.

Experimental study of how children learn to pronounce. 2, p. 182.

Analysis of silent reading progress with studies of the comparative difficulty of materials in different successive grades. 2, p. 146.

VI. Rating-scale Problems

Methods of discovery and encouragement of college students of superior ability and promise of achievement in research. 8, p. 439; 8, p. 239; 16, p. 254.

VII. Need for Statistical Devices and Methods

A method which permits study of the relationship of mental test scores and achievement in a simple direct and specific way. 3, p. 77.

A valid method for computing promotion rates. 5, p. 309.

In addition to the problems succinctly stated there are numerous evidences of the need of more adequate data to support conclusions and of a more careful definition of terms and usages.

It is often difficult to ascertain whether "diagnosis" refers to individual cases, to general class conditions or to larger units of organization.

"Analysis" is sometimes contemplated with reference to function and at other times to phases of end results, test scores, or pupil traits. The fact that the meaning attaching to these two terms is so often inadequately stated is, no doubt, one indication that new meanings are beginning to attach to their use.

Summary.—The three specific problems most frequently mentioned in the area to which this search was limited are:

Studies of the constancy of the IQ (5).[1]

Reformulation and analysis of silent reading (4).

Scientific determination of curriculum in social studies (4).

The fact that all of these problems are well represented in the titles of present contributions is an indication that they have already been attacked and are urgent because the need for further study is so frequently mentioned. No doubt the specific problems are more clearly defined as one result of pioneer studies in any field.

Frequency is not taken as a valid measure of significance in this connection. Pioneer studies in any field are the precursors of future trends. If statements of unsolved problems raised in the course of

[1]The total frequency of mention in the thirty-two magazines.

investigation were available in connection with the listed conclusions now so frequently found in reports, this exploratory function of pioneer studies would be more adequately served.

The classification of problems was not seriously hampered by overlapping and it is only necessary to count the references under each heading to get an idea of the relative frequencies of the problems thus grouped: Mental Tests 16; Curriculum Studies 15; Administration and Supervision 11; Educational Tests 9; Learning Studies 8; [illegible]ing Scales 3; Statistical Devices 2. We are led to wonder whether the paucity of studies in new statistical methods and devices is due to too great a reliance on some of those now so uncritically accepted, or whether some of our creative thinkers along this line are so deeply involved in carrying forward some specific line of research, that there is insufficient opportunity for the mental manipulation of relations which often leads to contributions of high order.

The leadership of educational writers, whose recent contributions were the material read in this search, gives added significance to the list of problems thus assembled.

A tabulation of the "Contents" of the same periodicals would give an interesting exhibit of the recent *results* of research, but that is not within the scope of this article.

In a recent article in the Journal of Applied Psychology, Terman gives quantitative data on the shift in the trend of psychological research, which has become increasingly noticeable since 1900. Of the researches in which the 306 members of the American Psychological Association are at present engaged, only 48.5 per cent are in pure psychology while 51.5 per cent are in applied fields. This development is in agreement with that in other sciences, and is an indication of the increasing "value" placed on psychology as a factor in the solution of human problems.

GROUP WILL-TEMPERAMENT TESTS

M. J. REAM

Bureau of Personnel Research, Carnegie Institute of Technology

In predicting success in school subjects, or success in specific vocations, the limitations of intelligence tests are recognized by their most enthusiastic advocates. The low correlations of test scores with success prove that there are other significant traits.

The Downey scale[1] of individual will-temperament tests is an attempt to bring to light some of these other factors. While individual testing presents many subtle observations not possible with group tests, yet a group method of giving the test was necessary if the Downey scale were to be included in a comprehensive study of successful and unsuccessful salesmen conducted at the Carnegie Institute of Technology. Accordingly, the Downey scale was modified at the Carnegie Bureau of Personnel Research so that groups of subjects might be tested at one time.

This series of group tests has been given during the past two years to 500 insurance salesmen, 600 Freshmen at the Carnegie Institute of Technology, and 150 stenographers, typists, and comptometer operators at a technical night school. The writer has carried on the evaluation of the test for the insurance salesmen only. Production records were available for about 125 of these salesmen. Some parts of the test have proved to be of positive value and are now included in the selection program for insurance salesmen prepared at the Carnegie Bureau of Personnel Research.

This group test, as here presented, follows the Downey scale rather closely in the test situations presented. Handwriting is used in eight of the eleven parts of the test, but none of the usual assumptions of graphology are made; only the changes in handwriting under the controlled experimentation of this test are considered in the results.

The group test differs from the Downey test in two or three important respects. First, the work-limit is changed to a time-limit basis, the necessity for which is obvious in group testing. Second, the giving of the test is less subjective, is less dependent on the examiner's personality and technique. The scoring of the test is objective and quantitative, which is essential if the test is to be used in the com-

[1] Downey, J. E.: The will-profile. A tentative scale for measurement of the volitional pattern. Department of Psychology, *Bull.* No. 3, University of Wyoming, 1919.

mercial field and handled by psychologically untrained examiners. Third, the group test is much shorter. It can be given in thirty minutes.

THE PARTS OF THE TEST

In describing the parts of the test, reference is made in each case to the name of the Downey individual test from which each part has been adapted. The name and the Downey definition appear in parentheses. Reference to these already familiar names will aid in identification. But whether these tests actually measure the traits indicated is not the concern of this article. Sufficient for the present purpose is the fact that *scores in the tests show a relationship* to successful sales work.

Parts 1 and 10. (*Speed of movement.* This test is intended to measure "normal speed of movement relative to size of person and age.")

The directions for Parts 1 and 10 of the group test are as follows:

In the space below, copy the words "United States" as you usually write them, in your usual style and at your usual speed. Copy the words repeatedly until you are told to stop. You do not need to hurry. Wait for the signal before beginning.

The time allowed is thirty seconds. One might expect considerable variation in speed of writing under different physical conditions. To counteract in a measure such variation, Part 1 is repeated as Part 10 near the end of the test. The score is the average number of letters written in the two parts.

Part 2. (*Motor inhibition.* The ability to retard writing is measured, which is related to "capacity to keep in mind a set purpose and achieve it slowly.")

The directions for Part 2 are:

Write each of the words as slowly as possible on the line after each word. Write the words AS SLOWLY AS YOU POSSIBLY CAN and still keep the pencil moving. Do not enlarge your writing. Wait for the signal before beginning.

This test is given three times. Two repetitions are necessary to make some subjects comprehend that writing just as slowly as possible is really wanted. Only the last repetition, for which sixty seconds are allowed, is scored. The score is the number of letters written.

Part 3. *Speed of decision in choosing better traits.* This test consists of an elaboration of the Downey list of opposite traits. Samples are "careful-careless," "slow-quick," "gloomy-cheerful."

The directions are:

Check the ONE trait in each pair which is the BETTER in most circumstances. For example, it is better to be careful than careless in most circumstances. So, careful should be checked in the first pair. Do not skip any pairs.

Forty-five seconds are allowed and the score is the total number of pairs of traits checked. The purpose of this part is merely to provide a basis for comparison with speed of decision in traits which apply to the subject personally. This second test of speed of decision appears later in the series.

Part 4. (*Freedom from inertia.* The ratio of the rate of normal writing to the rate of speeded writing is determined. This is intended to measure "quickness in warming up and tendency to work at one's highest speed without external pressure.")

The directions for Part 4 are: In the spaces below write the words "United States" *as quickly as you possibly can* and still have the writing legible. Continue until you are told to stop. Wait for the signal before beginning.

Sixty seconds are allowed. In scoring, the average number of letters written in Parts 1 and 10 above, is divided by the number of letters writen in Part 4. The resulting decimal is the score for Part 4.

Part 5. (*Motor impulsion.* Motor impulsion or the "tendency to impetuosity and energy of reaction is measured by the magnification and increased speed of writing under distraction.") In Part 5, visual control over writing is first eliminated by having subjects write the phrase "United States" without looking at the paper. This is repeated once. Next, the subjects must write the phrase repeatedly while counting the number of taps which the examiner makes on the table. The examiner diverts the attention of the subjects from the writing by tapping loudly and irregularly. The subjects must watch the examiner, not their papers. Two series of such writing under distraction are given, each lasting twenty-five seconds. The size of this writing and the speed, measured by the number of letters written, are compared with normal size and speed, measured in Parts 1 and 10. The resulting ratios added together give the score for this part.

Part 6*a*. *Success in disguise.* (Flexibility. Versatility in disguising one's handwriting is suggested as "characteristic of the histrionic or fluidic temperament.") An unpublished study of the students of dramatic art at the Carnegie Institute of Technology gives some evidence in support of this suggestion.

The directions for Part 6 are: Write the words "United States" in the space below trying to disguise your handwriting in as many ways as possible and as much as you can. Try out any disguise you can think of but do not print. Take as much time as you need and copy the words as many times as necessary. Keep trying until you feel that you have made a copy that even a handwriting expert could not identify as yours.

Three minutes are allowed for Part 6. The score is the number of different disguises accomplished. A sample scoring scale has been provided to facilitate the scoring of this test.

Part 6*b*. *Attempts at disguise.* (Volitional perseveration. This is defined as "persistence in attaining an indefinitely defined end.") Many subjects do not continue to work the entire time allowed for Part 6. Since the number of attempts at disguise varies, Part 6 is also used as the group modification of the individual test of perseveration which is the length of time taken for one disguise. Part 6 in this case is scored according to the number of attempts to disguise the writing of the phrase.

Part 7. (*Care for detail.* Accuracy in copying samples of handwriting is intended to measure "attention to details.") The directions for Part 7 are: Imitate the model sentences AS EXACTLY AS YOU CAN. Take as much time as you need. Wait for the signal before beginning.

Ninety seconds are allowed for this part. It is scored by a stencil in which inaccuracies are specifically indicated. Only two sentences are scored. The maximum number of errors is thirty.

Part 8. (*Coordination of impulses.* This part is a group form of the Downey test of writing in restricted space. It is intended to measure "capacity to handle a complex situation without forgetting either factor involved.")

The directions for Part 8 are: Copy each of the sentences below *as rapidly as possible* on the line after each sentence. Be careful not to let the writing extend beyond the end of the line. Remember, you are to write the sentences *as quickly as you possibly can.* Wait for the signal before beginning.

The sentences become increasingly difficult to compress in the limited spaces provided. Forty-five seconds are allowed. The score is the total number of letters written on the short lines. Letters which extend beyond the lines are not counted.

Part 9. (*Speed of decision.* The individual test is intended to

measure "quickness in reaching a decision or conclusion.") Part 9 presents the same list of paired traits as Part 3 but in this case there is a personal reference. The directions for Part 9 are: Check the ONE trait in each pair which describes YOU better. You may be in doubt in some cases; for example, you may be careful in some things and careless in others, but as a general rule you are more often one than the other. Check that trait. Do not skip any pairs. It will make no difference if you do not finish the whole list; speed does not count.

Sixty seconds are allowed and the score is the total number of pairs of traits checked. In this part, the decision called for is "subjective," involving self-judgments. A number of persons otherwise quite rapid in their reactions, show considerable blocking when checking these personal traits. The test is intended to measure ease and rapidity of decision in subjective items, and the tendency not to be critical. To throw this tendency into greater relief the score on this part is compared with the score made on Part 3 and a ratio computed. The score thus determined is treated as a measure of self-consciousness, on the thesis that the highly self-conscious individual will be proportionately slower in making subjective, personal judgments than in making non-personal decisions.

Part 11. (*Assurance.* This test is intended to measure the "degree of confidence with which one maintains his opinions against contradiction.") In this part a chart which contains Arabic and Roman numerals and small and capital letters, nine characters in all, is exposed to view for ninety seconds, after which the chart is withdrawn. The test blank contains fifteen statements about the chart, which are to be marked TRUE or FALSE, and doubly underlined if the subject feels especially sure of his answer. A final paragraph contradicts the actual conditions of the chart as follows: If you finish before time is called, you may check your accuracy in the last three statements. The word FALSE should be underlined after statements 13, 14, and 15. If you have not done this, you are at liberty to change your answers.

This suggestion is false as regards statements 14 and 15. The score on this part is based on the subject's resistance to suggestion and the number of answers which he doubly underlines. Part 11 has no time limit. The test papers are collected as soon as the subjects finish this part. It will be noticed from this description of the parts of the group test that all the items of the Downey individual test are retained in modified form except Resistance to Opposition, and Revision. An

added item is the ratio between the two checkings of traits, which is intended to show the effect of self-consciousness.

Relationship between Group Test and Downey Individual Test

In order to determine the correspondence between the scores and volitional patterns obtained from the group test and those obtained from the Downey individual test, a group of persons was tested, first with the group test and later tested individually with the Downey test. The subjects were twenty-one men (Juniors in the School of Industries at Carnegie) and two young women, making twenty-three subjects in all. The testing was done by four assistants who were trained in the method of giving the Downey individual test.

The scores on the group and individual parts of the test were correlated with the following results:

Table I

Parts 1 and 10, group test with Speed of movement, individual test	0.72
Part 2, group test with Motor inhibition, individual test	0.55**
Part 3, group test with Speed of decision, individual test	0.42
Part 4, ratio, group test with Freedom from inertia, individual test	0.05
Part 5:	
Magnification ratio, group test with Motor impulsion, individual test	0.54*
Speed ratio, group test with Motor impulsion, individual test	0.42*
Summed ratios, group test with Motor impulsion, individual test	0.50*
Part 6*a*, Success in disguise, group test with Flexibility, individual test	0.12*
Part 6*b*, Attempts, group test with Perseveration, individual test	0.90**
Part 7, group test with Care for detail, individual test	0.72*
Part 8, group test with Coordination of impulses, individual test	0.16*
Part 9, group test with Speed of decision, individual test	0.46
Part 11, group test with Assurance, individual test	0.42*

* Rank correlation formula used.

** Correlation ratio used, formula for non-linear distribution.

The results show relatively high correlations with those corresponding individual tests which are scored entirely objectively and quantitatively. These individual tests are: Speed of movement, motor inhibition, speed of decision, freedom from inertia, perseveration, and care for detail. The only low correlation in this group is that between Part 4—ratio and freedom from inertia. Scores in both of these tests are ratios derived from raw scores. Since ratios do not show a bell-shaped distribution, a high correlation is hardly to be expected.

Low positive correlations are found with those parts which are qualitatively scored or subjectively combined in the individual test—coordination of impulses, motor impulsion, assurance and flexibility.

The group and individual tests revealed volitional patterns of the same general type for each subject. There were occasional breaks in the correspondence but these may have been partly the result of repetition of certain parts in the second testing. This was particularly marked in the checking of traits. Two-thirds of the subjects raised their decile standings on this part in the later test, while only one subject lowered his decile standing. In view of all the results of this experiment it is evident that the group test is a fairly satisfactory approximation of the Downey individual test.

An incidental study was made of the test names and definitions published by Professor Downey. Scores and standings in the group test were reported to one group of thirty-five salesmen, together with

TABLE II.—REACTIONS OF THIRTY-FIVE SUBJECTS TO STANDINGS IN TRAITS. REPORTED FROM RESULTS OF GROUP WILL-TEMPERAMENT TESTS AND OTHER TESTS

Test	Per cent right	Per cent wrong	Per cent too high	Per cent too low
Parts 1 & 10 (Speed of movement)	69.0	31.0	14.0	17.0
Part 4 ratio (Freedom from inertia)	93.0	7.0	7.0	0.0
Part 9 (Speed of decision)	93.0	7.0	0.0	7.0
Part 6—Disguises (Flexibility)	90.0	10.0	3.0	7.0
Part 5 (Motor impulsion)	74.0	25.0	3.0	22.0
Part 11 (Assurance)	80.0	20.0	13.0	7.0
Part 9 ratio (Freedom from self-consciousness)	70.0	30.0	12.0	18.0
Part 2 (Motor inhibition)	88.0	12.0	9.0	3.0
Part 7 (Care for detail)	73.0	27.0	15.0	12.0
Average	81.1	18.8	8.4	10.3
Other tests:				
Intelligence	81.0	19.0	0.0	19.0
Meeting objections	84.0	16.0	3.0	13.0
Business information	73.0	27.0	9.0	18.0
Average	79.3	20.6	4.0	16.6

standings in other tests—intelligence, meeting objections and business information. Each person in the group was given a chart containing the Downey test names with his standings in the corresponding parts of the group test graphically presented. The Downey descriptive account of each test was read, after which each subject was asked to mark whether his standing indicated in the graph was correct, too high, or too low, according to his own judgment of himself concerning the various traits.

The results of these self ratings are shown in Table II.

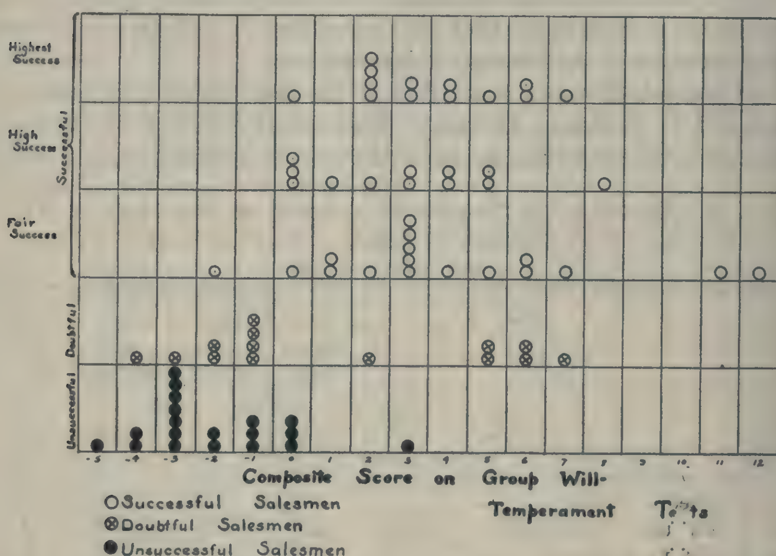


Fig. 1.—Predictive value of group will-temperament tests with forty-seven salesmen in first insurance school. X—Composite score in tests. Y—Success in selling insurance.

On the surface the results appear rather striking, but sources of error immediately suggest themselves. The high percentage marked "right" is without doubt partly the result of suggestion since many people have never attempted to analyze their own traits, much less to rate them. A self rating made before the test results are presented would be a better check. Nevertheless, it is worthy of note that the will-temperament tests received just as high a percentage of correct ratings as the non-volitional tests—intelligence, meeting objections, and business information.

RESULTS

The selling records of the salesmen were used as a criterion to determine the value of this series of group tests. During an eleven weeks course of instruction each salesman was required to spend each afternoon in actual field selling. The territorial as well as the time conditions were uniform. At the end of this period, standings were

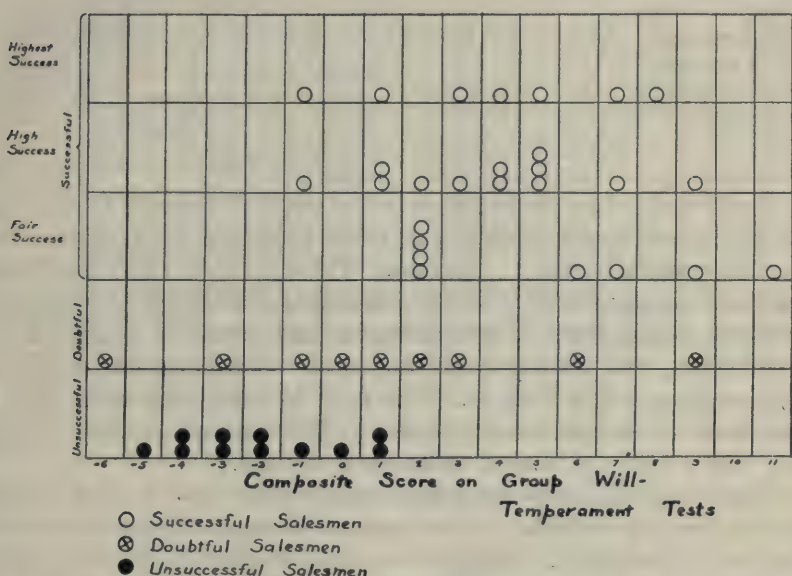


Fig. 2.—Predictive value of group will-temperament tests with seventy-five salesmen in second insurance school. X—Composite score in tests. Y—Success in selling insurance.

assigned on the basis of number of cases and amounts sold. The amount of insurance sold ranged from none at all to $140,000.00. The men were placed in five groups, from the entirely unsuccessful group to the most successful group. The tests were evaluated with this criterion.

The accompanying charts show the discriminating value of this series of tests with two separate groups of salesmen. Scores in the tests were statistically weighed and combined into a single composite score for each salesman. Median composite scores for each group are shown in the following table:

Test III

	Median composite score	
	Group A	Group B
Successful salesmen:		
Highest success	4	3
High success	4	3
Fair success	3	3
Doubtful salesmen	2	−1
Unsuccessful salesmen	−2	−3

Some tests showed upper or lower critical scores. Other tests showed discriminative value only when combined with another test in a three variable scatter diagram. The most useful tests for selecting salesmen are: Part 6 (success in disguise), Part 9 (speed in checking personal traits), Part 5 (magnification and speed of writing under distraction), and Part 9 ratio (speed of checking traits in Part 9 relative to speed in Part 3). The tests of least value are: Part 11 (resistance to contradiction), Parts 1 and 10 (normal speed of writing) and Part 4 (ratio of normal to speeded writing).

This series of group will-temperament tests proved itself of sufficient value to be included in the Bureau of Personnel Research selection program for salesmen. Results from these tests are much more significant for sales work than intelligence test results. The group tests are used in connection with an evaluation of previous training, experience, and a study of interests pertinent to sales work. A valuable instrument for use in vocational selection has resulted.

Conclusions

1. The series of group tests approximate fairly closely the results obtained from the Downey individual will-temperament test.

2. The tests are of positive value in predicting success in selling insurance.

COMPARATIVE VARIABILITY AT DIFFERENT AGES

V. A. C. HENMON
University of Wisconsin
AND
W. F. LIVINGSTON
Stoughton, Wisconsin

There is a widespread belief which frequently finds expression in the literature of education, that individual differences are greater during adolescence than at any other time in life and that the development from childhood to adolescence is not gradual but saltatory. G. Stanley Hall is the chief proponent of this doctrine. Concerning the adolescent period he says; "The human plant circumnutates in a wider and wider circle, and the endeavor should be to prevent it from prematurely finding a support, to prolong the period of variation to which this stage of life is sacred. . . ."[1] "The possibility of variation in the soul is now at its height."[2] "The forces of growth now strain to the uttermost against old restrictions. It is the age of bathmism, or more rapid variation, which is sometimes almost saltatory."[3] "Individual differences of all kinds are now suddenly augmented."[4] "The range of individual differences and average errors in all physical measurements and all psychic tests increases."[5]

This theory has important practical applications and its influence is plainly visible in our present systems of school organization. The contention that youth is the period of great fluctuation and that therefore throughout the high school age there is a decided increase in variability in all mental functions implies that the secondary school should provide a wider range of elections in the curriculum, smaller classes and more individualization of instruction, and greater versatility in methods of presentation of subject matter in order to appeal to the widely varying characteristics of a high school class. On the other hand (and this is the more serious consideration), the implied greater similarity between children in the grades offers an excuse for larger classes, for poorer teachers, and for forcing all pupils through

---

[1] Hall, G. Stanley: "*Adolescence*," Vol. II, p. 88.
[2] *ibid.*: p. 89.
[3] *ibid.*: p. 90.
[4] *ibid.*: p. 363.
[5] Hall, G. Stanley: "Youth," p. 6.

the same process and by the same methods until the approach of adolescence.

While thousands of children have been tested in a great variety of mental and physical traits, the writers have been unable to locate any systematic review of the available evidence on this supposedly greater variability during adolescence. It is a curious fact that in spite of the importance of the theory, no one appears to have taken the trouble to present the evidence for it, much less question its correctness. Common observation, confirmed by more exact study, shows the wider variability in height and weight at adolescence and, by analogy, the law appears to have been extended to mental traits without adequate investigation.

This study represents an examination of the comparative variabilities as revealed in some of the most representative studies of mental and physical development. Those investigations were used in which the number of cases was large, in which the variabilities had been determined, and in which norms for a wide range of ages were available.

The variabilities for different ages were rendered comparable by determining the coefficient of variation, obtained by dividing the measure of central tendency (average or median) by the measure of variability (average deviation, standard deviation or probable error). This is merely finding the per cent that the variability is of the central tendency from which the deviations were obtained. Whatever measures of central tendency or variability were used by the investigator were used for our calculations. While the essential problem was to compare variabilities at different ages, the variability in different grades is practically as important and this was determined in several sample cases. Incidentally, also, the ratios of female to male variability were calculated for the light they might throw on the mooted question of the variability of the sexes.

Variability in Physical Traits

The comparative variabilities in height, weight and lung capacity were computed from the data of Burk,[1] Baldwin,[2] and Gilbert.[3]

[1] Burk, F.: Growth of Children in Height and Weight, *American Journal of Psychology.* Vol. IX, 1898, pp. 253–326.

[2] Baldwin, Bird T.: Physical Growth and School Progress. *Bulletin* No. 10, U. S. Bureau of Education, 1914, p. 212.

[3] Gilbert, J. A.: Researches on the Physical and Mental Development of School Children. *Studies* from the Yale Psychological Laboratory, Vol. II, pp. 40–100, 1894.

Various other physical characteristics were studied also but are not reported here. In the interest of economizing space and to prevent confusion in examining the tables, only the total number of cases is given and not the number of cases at each age. The number of cases is in all cases sufficiently large to make the coefficients reliable. Table I gives the results for physical traits.

TABLE I.—COEFFICIENTS OF VARIABILITY AT DIFFERENT AGES IN PHYSICAL TRAITS

	Boys						Girls					
Ages	Height (1)	Height (2)	Height (3)	Weight (4)	Weight (5)	Lung capacity (6)	Height (1)	Height (2)	Height (3)	Weight (4)	Weight (5)	Lung capacity (6)
6.0	.....	0.032	0.034	0.093	0.098	0.169	.....	0.026	0.031	0.086	0.097	0.190
6.5	.....	0.036	....	0.098	.....	.....	.....	0.034	.....	0.087		
7.0	.....	0.042	0.033	0.108	0.091	0.169	.....	0.034	0.029	0.104	0.087	0.206
7.5	.....	0 038	...	0.123	.....	.....	.....	0.037	.....	0.086		
8.0	.....	0.040	0.035	0.122	0.112	0 155	...	0.033	0.035	0.098	0.096	0.159
8.5	0.043	0.034	.....	0.118	....	....	0.046	0.031	.....	0.098		
9.0	....	0.035	0.039	0.096	0.145	0.196	.....	0 037	0.035	0.108	0.116	0.135
9.5	0.044	0.039	.....	0.108	....	.....	0.044	0.033	....	0.113		
10.0	.....	0.035	0.032	0.122	0.100	0.168	.....	0.028	0.038	0.126	0.118	0.174
10.5	0.044	0.036	.....	0.107	.....	.....	0.046	0 032	.....	0.145		
11.0	.....	0.041	0.025	0.114	0.094	0.145	.....	0.032	0.040	0.101	0.085	0.125
11.5	0.044	0.034	.....	0.115	.....	...	0 048	0.038	.....	0.135		
12.0	.....	0.040	0.038	0.106	0.091	0.124	.....	0.039	0.042	0.160	0.134	0.142
12.5	0.046	0.040	.....	0.112	.....	.....	0.051	0.038	.....	0.145		
13.0	.....	0.039	0.038	0.142	0.106	0.146	.....	0.037	0.037	0.145	0.115	0.179
13.5	0.052	0.040	.....	0.143	.....	.....	0.048	0.040	.....	0.121		
14.0	.....	0.037	0.057	0.124	0.171	0.191	.....	0.035	0.045	0.123	0.135	0.165
14.5	0.055	0.044	.....	0.118	.....	.....	0.043	0.031	.....	0.136	.....	.....
15.0	.....	0.046	0.051	0.119	0.140	0.185	.....	0.032	0.034	0.100	0.100	0.133
15.5	0.054	0.047	.....	0.109	.....	.....	0.037	0.034	.....	0.103		
16.0	.....	0.039	0.032	0.117	0.093	0.161	.....	0.026	0.032	0.123	0.103	0.140
16.5	0.047	0.033	....	0.108	.....	.....	0.035	0.028	.....	0.091		
17.0	.....	0.031	0.018	0.077	0.087	0.163	.....	0.033	0.029	0.092	0.132	0.156
17.5	.....	0.024	.....	0.068	....	.....	.....	0.031	.....	0.083		
18.0	.....	0.032	.....	0.071	.....	.....	.....	0.030	.....	0.079		

(1) Burk's data, 88,449 cases.
(2) Baldwin's data, 1924 cases.
(3) Gilbert's data, about 1200 cases.
(4) Baldwin's data.
(5) Gilbert's data.
(6) Gilbert's data.

An examination of the table shows for height and weight approximate constancy in the coefficients for boys up to the age of 12½ years, with a sudden rise to the high point at 14 or 14½ years and a decrease thereafter. The highest points for the three sets of measurements for height are at 14½, 15½, and 14 years, respectively. For weight the high points are at 13½ and 14 years. The same general tendency holds for girls except that with them the highest point is reached roughly two years earlier. In height the highest points are at 12½, 13½ and 12 years, respectively. In weight the highest point is at 12 years in Baldwin's results and in Gilbert's data at 12 and 14 years. Both in height and weight, then, theory seems to hold for not only are the variabilities greatest at adolescence but the development is saltatory. Particular interest attaches to this table for height and weight are the only traits which we have been able to find where a large number of cases have been studied, in which the theory does hold, with one exception. The measurements of lung capacity, based on about fifty cases for each age, show no evidence of greater variability or saltation at adolescence.

Variability in Mental Traits

Coefficients of variability in mental traits were computed for the data reported by Gilbert,[1] Pyle,[2] and Bickersteth.[3] In Gilbert's results there are about fifty cases at each age. In Pyle's norms the number of cases varies widely. Where the number involved is very small, the figures in Tables III and IV are enclosed in parentheses.

Table II gives the results of the computations for the eight mental tests in Gilbert's research with the averages for all the tests for the ages from six to seventeen years.

Tables III, IV and V give similarly the coefficients for the data of Pyle and Bickersteth.

A detailed examination of these tables shows that the period of greatest variability in mental traits is during the years of childhood, not at adolescence. The coefficients tend to decrease with fair uniformity from childhood to adulthood. It is most strikingly shown in Pyle's test of invention but holds almost equally well in the opposites,

---

[1] *Op. cit.*

[2] Pyle, W. H.: "The Mental Examination of School Children," New York, 1913, p. 70.

[3] Bickersteth, M. E.: The Application of Mental Tests to Children of Various Ages, *British Jour. of Psychol.*, Vol. IX, Dec., 1917.

TABLE II.—COEFFICIENTS OF VARIABILITY AT DIFFERENT AGES IN EIGHT MENTAL TESTS (GILBERT'S DATA)

Boys

Test ages	6	7	8	9	10	11	12	13	14	15	16	17
Time memory	0.444	0.324	0.460	0.471	0.449	0.409	0.609	0.893	0.450	0.443	0.434	0.406
Reaction time	0.163	0.172	0.155	0.222	0.124	0.167	0.152	0.163	0.167	0.137	0.109	0.129
Reaction time (Disc. & Choice)	0.099	0.180	0.119	0.142	0.122	0.150	0.156	0.142	0.122	0.177	0.124	0.115
Force of suggestion	0.391	0.356	0.300	0.210	0.312	0.303	0.237	0.241	0.306	0.318	0.312	0.480
Fatigue	0.412	0.431	0.336	0.297	0.343	0.320	0.333	0.424	0.348	0.355	0.300	0.434
Voluntary motor ability	0.119	0.118	0.136	0.097	0.094	0.108	0.102	0.101	0.102	0.083	0.091	0.068
Muscle sense	0.400	0.333	0.377	0.431	0.511	0.372	0.397	0.500	0.576	0.355	0.400	0.433
Sensitiveness to color differences	0.216	0.253	0.240	0.360	0.316	0.283	0.312	0.327	0.291	0.268	0.300	0.350
Averages	0.280	0.271	0.265	0.278	0.284	0.264	0.287	0.348	0.283	0.279	0.258	0.302

Girls

Test ages	6	7	8	9	10	11	12	13	14	15	16	17
Time memory	0.346	0.298	0.391	0.262	0.340	0.493	0.498	0.542	0.574	0.561	0.285	0.445
Reaction time	0.183	0.165	0.119	0.192	0.191	0.165	0.177	0.170	0.160	0.143	0.151	0.159
Reaction time (Disc. & Choice)	0.127	0.180	0.116	0.157	0.108	0.150	0.132	0.133	0.152	0.110	0.111	0.140
Force of suggestion	0.400	0.356	0.273	0.212	0.284	0.288	0.220	0.237	0.283	0.276	0.260	0.387
Fatigue	0.328	0.331	0.304	0.376	0.374	0.305	0.478	0.394	0.508	0.496	0.479	0.318
Voluntary motor ability	0.127	0.118	0.092	0.116	0.104	0.108	0.101	0.117	0.121	0.108	0.107	0.073
Muscle sense	0.309	0.333	0.418	0.440	0.478	0.500	0.394	0.535	0.416	0.305	0.353	0.406
Sensitiveness to color differences	0.187	0.219	0.328	0.333	0.365	0.347	0.294	0.414	0.304	0.239	0.325	0.286
Averages	0.251	0.250	0.255	0.261	0.280	0.294	0.287	0.318	0.314	0.279	0.259	0.278

genus-species, and part-whole tests. The general tendency is clearly revealed in Table VI, which summarizes in a single table the variability in mental traits, combining the results of Pyle, Bickersteth and a portion of those of Gilbert, and disregarding sex differences.

In all of Pyle's data, the only evidence for greater variability and saltation is found in the free association test. In Gilbert's results, the time memory test and the fatigue test (in girls only) are the only ones that support the theory. The results in Gilbert's force of suggestion test are peculiar in that the variabilities are greatest at the extreme age limits, six and seventeen years.

Considerable interest attaches to the variability in a general intelligence test in view of Freeman's article in this journal on the assumptions underlying the calculation of intelligence quotients with group

TABLE III.—COEFFICIENTS OF VARIABILITY FOR DIFFERENT AGES IN ELEVEN MENTAL TESTS (PYLE'S NORMS)

Boys

Test ages	6	7	8	9	10	11	12	13	14	15	16	17	18	Adults
Logical memory	...	...	0.271	0.317	0.223	0.170	0.210	0.171	0.193	0.183	0.162	0.251	0.162	0.182
Rote memory (abs.)	...	...	0.340	0.285	0.261	0.223	0.212	0.237	0.190	0.181	0.207	(0.150)	(0.054)	0.147
Rote memory (con.)	...	...	0.214	0.228	0.173	0.169	0.132	0.146	0.160	0.121	0.145	(0.111)	(0.155)	0.149
Digit-symbol	...	...	0.339	0.325	0.253	0.220	0.267	0.256	0.214	0.186	0.217	0.180	0.121	0.292
Symbol-digit	...	...	0.530	0.378	0.310	0.280	0.275	0.231	0.229	0.227	0.283	0.182	0.282	0.275
Word building	...	...	0.600	0.493	0.329	0.345	0.372	0.395	0.364	0.275	0.219	0.158	0.286	0.141
Free association	...	...	0.326	0.278	0.303	0.342	0.318	0.430	0.390	0.370	0.438	0.280	0.339	0.327
Opposites	...	...	0.366	0.357	0.413	0.266	0.252	0.310	0.296	0.325	0.285	0.168	0.143	0.149
Genus species	...	...	0.739	0.596	0.569	0.458	0.352	0.380	0.361	0.486	0.282	0.292	0.346	0.264
Part—whole	...	...	0.654	0.446	0.342	0.314	0.381	0.387	0.336	0.371	0.333	0.253	0.290	0.194
Cancellation	...	...	0.294	0.250	0.269	0.263	0.275	0.212	0.248	0.243	0.217	0.267	0.209	0.198
Averages	...	...	0.425	0.359	0.313	0.277	0.277	0.287	0.271	0.270	0.253	0.226	0.242	0.211

Pyle, W. H.: "Mental Examination of School Children."

TABLE IV.—COEFFICIENTS OF VARIABILITY FOR DIFFERENT AGES IN ELEVEN MENTAL TESTS
(PYLE'S NORMS)
Girls

Test ages	6	7	8	9	10	11	12	13	14	15	16	17	18	Adults
Logical memory	...	...	0.396	0.303	0.203	0.211	0.186	0.184	0.192	0.161	0.136	0.188	0.116	0.147
Rote memory (abs.)	...	...	0.380	0.241	0.196	0.223	0.220	0.111	0.215	0.193	0.127	(0.195)	(0.081)	0.153
Rote memory (con.)	...	...	0.215	0.189	0.131	0.137	0.157	0.133	0.158	0.166	0.115	(0.113)	(0.038)	0.162
Digit-symbol	...	...	0.245	0.261	0.234	0.221	0.215	0.222	0.186	0.171	0.192	0.200	0.270	0.130
Symbol-digit	...	...	0.486	0.325	0.329	0.321	0.286	0.250	0.222	0.222	0.182	0.197	0.141	0.172
Wordbuilding	...	...	0.594	0.535	0.394	0.327	0.380	0.335	0.362	0.198	0.296	0.323	0.188	0.163
Free association	...	...	0.346	0.287	0.335	0.328	0.420	0.438	0.329	0.343	0.344	0.336	0.297	0.341
Opposites	...	...	0.500	0.381	0.284	0.267	0.259	0.289	0.224	0.294	0.217	0.156	0.132	0.170
Genus species	...	...	0.654	0.462	0.410	0.451	0.311	0.336	0.271	0.300	0.329	0.306	0.289	0.245
Part—whole	...	...	0.563	0.406	0.371	0.350	0.370	0.324	0.256	0.321	0.266	0.296	0.233	0.172
Cancellation	...	...	0.280	0.283	0.280	0.255	0.234	0.219	0.230	0.222	0.220	0.227	0.160	0.177
Averages	...	...	0.423	0.334	0.288	0.281	0.276	0.258	0.241	0.235	0.220	0.250	0.203	0.185

Table V.—Coefficients of Variability at Different Ages in Thirteen Mental Tests
(Bickersteth's Data)
Girls

Test ages	7	8	9	10	11	12	13	14	15
Number test (1)	0.575	0.443	0.375	0.346	0.398	0.371	0.345	0.349	0.288
Number test (2)	0.397	0.399	0.242	0.308	0.301	0.354	0.298	0.279	0.256
Alphabet test (1)	0.947	0.530	0.382	0.438	0.404	0.327	0.327	0.305	0.348
Alphabet test (2)	0.522	0.383	.280	0.273	0.319	0.340	0.327	0.265	0.265
Precision and speed of movement	0.048	0.052	0.056	0.071	0.076	0.094	0.106	0.151	0.116
Rate of tapping (1)	0.124	0.121	0.084	0.087	0.077	0.069	0.070	0.099	0.103
Rate of tapping (2)	0.309	0.397	0.332	0.386	0.572	0.379	0.483	0.468	0.521
Sustained attention test	0.320	0.235	0.218	0.297	0.279	0.219	0.218	0.203	0.192
Divided attention test	0.327	0.372	0.350	0.282	0.362	0.376	0.360	0.398	0.240
Averages	0.396	0.314	0.257	0.276	0.309	0.281	0.281	0.279	0.269
Memory for narrative	.....	.....	0.213	0.193	0.172	0.126	0.123	0.176	0.167
Memory for related words	.....	.....	0.158	0.210	0.187	0.246	0.178	0.163	
Spot-pattern test	.....	.....	0.233	0.184	0.153	0.213	0.182	0.177	0.219
Analogy test	.....	.....	0.212	0.253	0.390	0.191	0.186	0.180	0.168
Averages	.....	.....	0.204	0.210	0.225	0.194	0.167	0.174	0.185

(*British Journal of Psychology*, December, 1917—Bickersteth, M. E.)

Table VI.—Summary of Variability in Mental Traits

Age	Number of cases	Coefficient of variability
7	408	0.417
8	1324	0.356
9	2064	0.313
10	2387	0.283
11	2184	0.280
12	2466	0.266
13	2345	0.262
14	1906	0.253
15	1260	0.257
16	830	0.253
17	606	0.266
18	428	0.206
Adults	1570	0.205

test results. As Freeman points out, there are either or both of two assumptions involved in such calculations, *viz.*, decreasing rate of growth in mental development or diverging lines of growth. Table

VII gives the coefficients of variation calculated from Mrs. Pressey's data.[1]

TABLE VII.—COEFFICIENTS OF VARIABILITY IN PRESSEY'S GROUP INTELLIGENCE TEST

Age	Boys	Girls
8	0.256	0.150
9	0.254	0 211
10	0.258	0.203
11	0.245	0.145
12	0.168	0.132
13	0.142	0.117
14	0.138	0.098
15	0.141	0.088
16	0.081	0.080

There is here no evidence whatsoever of increasing variability or saltation. On the contrary, the decrease in variability is as unmistakable as it has been shown to be in practically all special mental traits.

VARIABILITY IN MENTAL TRAITS BY GRADES

Even if there is no evidence for increasing variability and saltation with age, it might still be urged that, after all, pupils are not classified in school on the basis of age and that on a classification according to grade, the wider range of differences at adolescence might reveal itself. Many such measurements have been made but only two will be reported here. The first are the coefficients for five of the Courtis tests given to 27, 171 children in the New York School Survey. Table VIII gives the facts.

The second are the variabilities in Language Scale A of the Trabue Completion Tests for which results are available for a large number of cases from Grade II upward.[2] Table IX gives the coefficients of variability for these data.

---

[1] Pressey, Luella W.: Sex Differences Shown by 2544 SchoolChildren. *Jour. of Applied Psychol.*, Vol. II, Dec., 1918, pp. 323–340.

[2] Trabue, M. R.: Completion Test Language Scales. Columbia Univ., *Contrib. to Educ.*, 1916.

TABLE VIII.—COEFFICIENTS OF VARIATION IN COURTIS ARITHMETIC TESTS (NEW YORK SURVEY DATA)

Grade	Test 1	Test 2	Test 3	Test 4	Test 5	Average
4	0.166	0.299	0.232	0.242	0.199	0.228
5	0.207	0.194	0.242	0.242	0.187	0.214
6	0.168	0.231	0.249	0.150	0.165	0.193
7	0.147	0.135	0.235	0.181	0.159	0.171
8	0.109	0.111	0.123	0.203	0.145	0.138
9	0.162	0.185	0.188	0.194	0.154	0.177
10	0.173	0.141	0.195	0.191	0.143	0.169
11	0.159	0.128	0.188	0.176	0.144	0.159
12	0.167	0.119	0.175	0.192	0.168	0.164

TABLE IX

Grade	Number of cases	Coefficient of variability
II	1318	0.454
III	1437	0.380
IV	1463	0.290
V	1507	0.196
VI	1454	0 165
VII	1456	0.148
VIII	1427	0.144
IX	273	0.140
X	171	0.116
XI	136	0.094
XII	103	0.103
College graduates........	114	0.067

The results show a rapid decrease up to the fifth grade and a gradual decrease thereafter.

DISCUSSION AND CONCLUSIONS

It is very evident that the law of increasing variability at adolescence does not hold for mental traits, so far as the groups for which measurements are available are concerned. On the contrary, there is in the school groups a marked reduction in variability at adolescence as contrasted with childhood. How is this reduction to be accounted for, particularly in view of the results of experiment on the effects of equal practice on individual differences, which have uniformly shown that differences do not decrease but rather increase when opportunities

for practice are equalized? In a certain sense it may be true that the range of differences is greater at adolescence if we include at each age the mentally deficient whose abilities in any test would be zero. School children from whom norms in mental tests are usually obtained are a selected group. Even so, in view of the large reduction in the coefficients, it is pretty certain that the average variability would not show an increase with age, provided a proportionate number of borderline and feebleminded children were tested and these results included in the distributions. In any case, the pedagogical inferences are based on the normal school population. Selection by eliminating those at the lower end of the distribution curve accounts, then, in part for the reduced variability found but not for all of it.

Inadequacy of training causes a narrowing of the distribution at the upper end. It has been shown over and over again that under proper stimulation, a very great increase in efficiency in mental functions is obtainable even in those traits which in the ordinary circumstances of life, are much practiced. In other words, there are possibilities of very great increases in efficiency in the upper ranges which are not realized and are not revealed in the test norms actually obtained. The norms, for example, for the Courtis Tests are considerably lower than they would be if the stimulus of experimental conditions were provided.[1] There is then no contradiction between these findings and the experiments on the effects of equal practice on individual differences. Under ordinary conditions, the effects of equalizing practice is to reduce individual differences since a certain modicum of efficiency is all that is required. When a sufficient stimulus is provided, the upper limit is greatly extended and both the range and average variabilities are greatly increased. There is then a possibility that individual differences may increase at adolescence but there is no evidence that they actually do.

What we need for a final answer to the problem is repeated measurements of a great number of unselected individuals over the entire period of childhood and adolescence. Such data, are of course, nowhere to be found now.

Sex Differences in Variability

Incidentally, in connection with this study, the ratios of the vari-

[1] Henmon, V. A. C.: Improvement in School Subjects Throughout the Year. *Jour. of Educ. Research*, March, 1920.

TABLE X.—RATIOS OF VARIABILITY OF GIRLS TO THAT OF BOYS

Ages	6	7	8	9	10	11	12	13	14	15	16	17	18	Adults	Av.
Physical traits	0.953	0.965	0.902	0.916	1.047	0.915	1.241	1.051	0.876	0.725	0.939	1.267	.....	....	0.983
Mental traits (Pyle)	.....	.....	1.031	0.936	0.931	1.016	1.003	0.905	0.908	0.924	0.877	1.136	0.914	0.914	0.958
Mental traits (Gilbert)	0.964	0.939	0.945	0.992	1.047	1.083	1.009	0.982	1.102	1.028	1.063	0.989	.....	.....	1.012
Pressey group intelligence test	.....	.....	0.586	0.831	0.787	0.592	0.786	0.824	0.710	0.624	0.988	.....	.....	.....	0.747

ability of girls to that of boys were computed. A summary of the results appears in Table X. The physical traits are height, weight and lung capacity, the data being those reported in Table I. The mental traits are those involved in the eleven tests by Pyle, the eight tests by Gilbert, and the group intelligence test by Pressey. The ratios for Pyle's and Gilbert's data were calculated for each test at each age and then averaged. They are not the ratios for the averages of the coefficients.

While the results in general seem to show a greater variability among the boys, the differences are not great, except in the Pressey Test, and there are marked irregularities, notably at seventeen years. There are large discrepancies between the data of Pyle and Gilbert at eight, ten, fourteen, fifteen and sixteen years. The Pressey data show a remarkably greater variability among the boys, far larger than any other the writers have been able to discover.

IS THE RATING OF HUMAN CHARACTER PRACTICABLE?

HAROLD RUGG

The Lincoln School of Teachers College

Agreement in Numerical Rating is not an Index of Agreement in Judgment of Character

I have illustrated by a striking exception the great difficulty—almost impossibility—of securing agreement in judging character. The isolation of this case will be made more evident by an accumulation of cases in which the details of man-to-man comparisons are reviewed. The unordered—yes, the chaotic—character of the judgments appears, irrespective of what traits are considered or of what kinds of scales are compared. I now believe that the evidence establishes the futility of obtaining single "ratings" on point scales of such dynamic qualities as "intelligence," "personal qualities," "general value to the service," "leadership," "physical qualities," "teamwork," and the like. The cases to be presented will show: (1) scales similar at both extremes but widely divergent in the middle; (2) scales alike at the lower end but dissimilar throughout the rest of the range; (3) scales in fair agreement but ratings made against them in great disagreement; (4) scales lacking in equivalence but ratings made against them in close agreement; (5) exact agreements in comparing one man with another paralleled by large disagreement in comparing him with a third; etc.

I take as the first illustration judgments of "physical qualities" and of "intelligence." The scales of Nos. 37 and 38 and the ratings which were made against them are reproduced in Table VIII. Here is an instance in which Nos. 37 and 38 used the same man at "15" and the same man at "3." The scales are alike at the extreme ends. Furthermore, the same man who appears on No. 37's "physical" scale at "12," also appears at the same value on No. 38's "leadership" and "intelligence" scales. Hence, the two physical scales probably represent closely the same differentiation. The same five men were rated against the physical scales. In only one instance was there close agreement in judgment. Nos. 37 and 38 agree that Staker is about the poorest captain, physically, they have known. No. 37 judges No. 4 to be as poor as Staker, while No. 38 rates No. 4, 6 points higher. Similarly, No. 38 rates No. 11, 2 points higher than does No.

TABLE VIII.—COMPARISON OF SCALES CONSTRUCTED BY NO. 37 AND NO. 38 TOGETHER WITH THEIR RATINGS ON SAME OFFICERS

Average position on others scales	Average of conf. ratings on him	No. 37's scales		Ratings assigned by No. 37 to officers rated by both No. 37 and No. 38	No. 38's scale		Ratings assigned by No. 38 to officers rated by both No. 37 and No. 38
		Values	Number or name of scale officer		Values	Number or name of scale officer	
Physical Qualities							
		15	Bradley	....	15	Bradley	
13.5	77	12	No. 36	Nos. 17, 27	12	Eggleston	
							No. 27
		9	Willard	No. 30	9	No. 35	Nos. 4, 30
13.5	67						Nos. 17, 11
		6	Holzinge	No. 11	6	No. 37	
9.0							
		3	Staker	No. 4	3	Staker	
Intelligence							
		15	Underwood	....	15	Luskin	
		12	Elwood	No. 11	12	No. 36	No. 17
9.8	77						
9.8	77	9	No. 36	Nos. 17, 27	9	Elwood	Nos. 11, 27
		6	Ballinger	No. 30	6	No. 35	No. 30
12.0	67						
		3	Willard	No. 4	3	Holzinge	No. 8

37. There is no general tendency for No. 38 to rate higher than No. 37, however, for while they agree on Bradley at the highest end of their scales, No. 37 rates No. 17 four points higher than does No. 38, and 2 points higher for No. 27. At the same time they agree on No. 30, each giving him 9 points. The topsy-turvy character of the ratings is brought out by such examples as these.

The "intelligence" scales of Nos. 37 and 38 provide quite a different sort of comparison—namely, a case in which ratings are made against scales that do not represent equivalent amounts of the trait.

See Table VIII. The instance is a clear exposition of the difficulty that is encountered in discriminating men who appear near the middle of a scale. Our tables show that there is a much larger probability that the "15" and the "3" scale-men will be more adequately discriminated than the "6," "9," and "12" men. It should be remembered that these scales were constructed by arranging the original lists in specific rank order. Thus there must be at least 4, 5, or 6 men represented between Elwood and No. 36, who are reversed on the two scales, occupying the "12" and "9" positions on No. 37's scale and the corresponding "9" and "12" positions on No. 38's scale. We have no means of stating the qualifications of the men who must have separated Elwood and No. 36 in these two original lists but certainly the conclusion can be drawn that there must have been a wide discrepancy in estimating the intelligence of the men rated. The ratings on them are:

Rating of Intelligence of 5 Persons

	No. 4	No. 11	No. 17	No. 27	No. 30
No. 37's ratings	3	12	9	9	6
No. 38's ratings	8	9	12	9	6

Two cases occur out of five in which there is exact agreement in total ratings built upon scales that reverse the scale-men, against whom the particular judgments must have been made. At the same time two other men are rated 12–9, 9–12 against these very same scale-men, and in the case of No. 4, there is a difference of 5 points, or one-third of the total scale. Instability of judgment, lack of assurance that the score represents equivalent merit, the influence of particular qualities on final judgment, these conclusions and suggestions occur to one as a result of studying such figures.

Before leaving this part of the discussion let us make one more comparison of scaling and rating "intelligence." Table IX supplies the data, together with intelligence test scores and average-ratings on each man. Note that No. 21 is rated one interval below No. 17 on No. 7's scale, but three intervals below No. 17 on No. 22's scale. Furthermore, in the construction of the scales McKinley is two intervals *superior* to No. 17 on No. 7's scales, whereas he is one interval *inferior* on No. 22's scale—a difference of 9 points or three-fourths of

the total scale. Such evidence shows that it is exceedingly difficult to maintain even the same rank order in placing men on the scale and in rating others against them. Note, too, that No. 43 is given "15" on No. 7's scale and "8" on No. 22's scale. This results in some interesting anomalies. No. 7 judges him to be as intelligent as McKinley and two intervals (6 points) better than No. 17. No. 22 judges him to be 4 points inferior in intelligence to McKinley and 7 points inferior to No. 17. Contrasted with these dissimilarities in rating, note that No. 24 is rated 9 on each scale and No. 21, 6 on each scale. The table proves, however, that *we may not deduce, from the fact that an officer is given exactly the same rating by two officers, that the rating represents closely similar estimates of intelligence as contributed to by "man-to-man" comparison. Analysis of such cases, which I am certain are typical, shows that rating scales made even under such well-controlled conditions as were those at Camp Taylor, will contain discrepancies in placing scale-men and in estimating human traits upon them, of between one and two scale intervals—that is between* 25 *and* 50 *per cent of the total scale.*

TABLE IX.—COMPARISON OF SCALES CONSTRUCTED BY NO. I AND NO. 22 TOGETHER WITH THEIR RATINGS ON SAME OFFICERS

Average position on others, scales	Average of conf. ratings on him	No. 7's scale		Ratings assigned by No. 7 to officers rated by both No. 7 and No. 22	No. 22's scale		Ratings assigned by No. 22 to officers rated by both No. 7 and No. 22
		Values	Number or name of scale officers		Values	Number or name of scale officers	
			Intelligence				
		15	McKinley	No. 43	15	No. 17	
10.8	82.0						
4.3	51.8	12	No. 4	No. 4	12	McKinley	
10.8	82.0	9	No. 17	No. 24	9	No. 7	Nos. 4. 11, 24
11.0	51.8	..	....	....	..	....	No. 43
		6	Whitfield	No. 21, 11	6	No. 21	No. 21
4.5	56.0						
3.0	55.5	3	No. 28	....	3	No. 32	
6.0	63.0						

We turn next to some illustrations of ratings on a person's general qualities. In making the army rating scale, the practical army officers insisted on adding a group of qualities called "general value to the service." In addition to judging a man's intelligence, his personal qualities, his physical qualities, and his ability as a leader, they wished to measure what he was worth to the army as an all-round man. Hence, we have, in scales for "general value," a summary evaluation much like that obtained by totalling the estimates of particular traits. However, there is no discernible difference in accuracy or inaccuracy in rating such a totality as distinguished from rating a more particularized group of qualities.

Two sets of scales are given in Tables X and XI. The scales of Nos. 11 and 19 provide a very helpful comparison of scale-placement

TABLE X.—COMPARISON OF SCALES CONSTRUCTED BY NO. 19 AND NO. 11 TOGETHER WITH THEIR RATINGS ON SAME OFFICERS

Average position on others scales	Average of conf. ratings on him	No. 19's scale: Values	No. 19's scale: Number or name of scale officer	Ratings assigned by No. 19 to officers rated by both No. 19 and No. 11	No. 11's scale: Values	No. 11's scale: Number or name of scale officer	Ratings assigned by No. 11 to officers rated by both No. 19 and No. 11
General Value							
		40	McKinley	....	40	No. 17	No. 37
30.4	82.0						
		32	Hotze	No. 11	32	No. 12	No. 12
32.0	70.0						
				No. 37			
							No. 21
30.4	82.0	24	No. 17	No. 24 Staker No. 12 No. 21 No. 22	24	Rumpel	Nos. 24, 19
22.0	55.6	16	No. 7	....	16	No. 7, Staker	No. 22
22.0	55.6						
11.4	51.8	8	No. 4	....	8	No. 4	
11.4	51.8						

Table XI.—Comparison of Scales Constructed by No. 22 and No. 19 Together with Their Ratings on Same Officers

Average position on others, scales	Average of conf. ratings on him	No. 22's scale		Ratings assigned by No. 22 to officers rated by both No. 22 and No. 19	No. 19's scale		Ratings assigned by No. 19 to officers rated by both No. 22 and No. 19
		Values	Number or name of scale officer		Values	Number or name of scale officers	
General Value							
		40	McKinley	....	40	McKinley	
							No. 11
30.4	82.0	32	No. 17	No. 11	32	Hotze	
							No. 24
24.0	63.0	24	No. 23	....	24	No. 17	
30.4	82.0						
				No. 24			
							No. 21
11.4	51.8	16	No. 4	....	16	No. 7	
22.0	55.6						
14.0	56.0	8	No. 21	No. 21	8	No. 4	
11.4	51.8						

and ratings for "general value" because of the fact that 3 of the 5 scale-men are the same on the two scales. Furthermore, four officers have been rated against these scale-men. Note that the two scales are equivalent at the low end but that No. 11's scale contains No. 17 at "highest," whereas No. 19's scale places No. 17 half way down the scale, at "middle." Here is an instance of wide disagreement (6 points) in the placing of one of the scale-men used, with perfect agreement in placing two more. The suggestion will occur that it might be caused by the difference in the "spread" of ability represented in the acquaintance of the two men. It probably is not, however, for McKinley, No. 19's "highest" man, is used by No. 11 at "highest" on intelligence; and No. 12, who appears as "high" man on No. 11's scale, is used twice as "high" man on 19's scale. Thus two of these three men are known to both No. 19 and No. 11 and there is a definite tendency to agree on the placement of these men in other qualities. Hence, the lack of agreement in scaling No. 17 must be due to distinct differences in estimating the abilities of the two men.

Now let us compare the ratings on these scales which are obviously not equivalent above the "6" point. No. 19 gives No. 21, No. 12, No. 24 and No. 22 very closely the same rating—namely, 22, 23, 24 and 20 respectively. But, No. 11 rates No. 12 as twice as valuable to the service (32 points) as No. 22, who gives him 16 points! In the same fashion No. 11 rates No. 12 one whole interval on the scale better than No. 24 at the same time that No. 19 judges him to be slightly poorer than No. 24.

No. 19 rates No. 21 three points lower than No. 17, whereas No. 11 rates No. 21 fourteen points lower. In this instance a difference in judgment in placing the scale-men merely accentuates the difference in judgment in rating on the scale.

On the other hand, No. 37, a major, is rated "30" by No. 19, that is, 7 points superior to No. 12. At the same time No. 37 is rated "40" by No. 11, that is, 8 points superior to No. 12, who appears on No. 11's scale at "32." In this case the rating on "general value," which differs by 10 points, represents closely the same relative judgment of two men who were involved in the comparison. It is clear that when the two scale-men at the lowest end of the scales are the same it does not necessarily follow that judgments made near the middle of the scale will be closely the same. No. 22 is rated "20" and "16" respectively by the two raters when compared with No. 7 and No. 4, who appear at the two lowest points on the scale. The difference in placement of No. 17 has contributed to very material differences in rating at the high end of the scale. Another instance of wide lack of agreement in judgment is found in the rating of Staker, a major, who is given "24" by No. 19 and "16" by No. 11. Furthermore, in direct man-to-man comparison he is rated "8," that is, as equal to No. 7 by No. 11 and 8 points better than No. 7 by No. 19. If such a divergence appears small it should be remembered that a similar difference in rating on all other qualities of the scale will amount, on the average, to a difference of 20 points in total rating.

Do not such illustrations[1] raise serious doubts concerning the validity of ratings of human traits on point scales? They prove to me that the task of comparing one person's qualities with another's is fraught with so much difficulty as to be impractical in rating the rank and file of persons and for most practical activities of life. This

[1] I omit many other illustrative tables and scales because of lack of space The situation for "personal qualities" and "leadership" is precisely the same as for those reported.

study is convincing of the difference in distinguishing persons at the extreme ends and the middle portion of the scale. If a person stands out conspicuously from his group for the presence or lack of a particular quality, it is much easier for his associates to agree in discriminating that quality.

But this very fact brings to the forefront one of the most important characteristics of the process of judging character. That is the role played by conspicuous traits in dominating reactions to total personalities.

How Do We Judge Our Fellows?

The Dominating Role of General Mental Attitudes and of Conspicuous Traits.—With considerable hesitation I advance, at this point, a theory to help explain the process of judging human character. I shall merely outline it at this time, wishing to elaborate it more fully later:

Two facts seem to be of paramount significance: first, we rate or judge our fellows in terms of a general mental attitude toward them; second, there is dominating this mental attitude toward the personality as a whole, a like mental attitude toward particular qualities. Some illustrations will supply the basis for these statements.

The striking case of Captain X.—Take first the most objectified case we have, a case in which separate judgments of a person's intelligence can be compared directly with several objective measures of his intelligence. Captain X was so well known and *was so conspicuous in his group* that he was used by 13 officers on 20 different subordinate scales—physical qualities, intelligence, leadership, etc. On each of these 20 scales he was elected to be "the poorest man I ever knew." Furthermore, he was so very conspicuous that three officers used Captain X as the "3" (lowest) man on four out of five of their scales. To them he was so outstandingly a weak man that there was no question of using another fellow captain for the lowest position on the different scales.

Now consider the objective measures of his abilities. *On three different psychological tests* (written group tests), *Captain X was first ranking man among* 151 *officers.* He scored 206 out of a possible 212 in the Army Alpha test. He scored 151 and 144 respectively on two forms of the Thorndike Alertness Test (which is Part I of his college entrance examination). He completed the test each time within the time limit of 30 minutes—29 minutes and 20 minutes respectively.

Moreover, he had been regarded only a few years before as an all-round man for he was a Rhodes Scholar at Oxford from a middle-western state university. At Oxford he made such a record that he was excused from certain examinations. *Here then is a startling example of divergence between ability-to-do and our judgment of it.*

Now, what was the explanation? I asked, separately, 8 of those who used him on the scale, why they had used him at "3." Their comments pointed out, indubitably, that their estimates of Captain X's intelligence, his physical qualities, his leadership, were dominated by their opinions of his personal qualities. They were unanimous in saying that it was impossible to "live with him." He was a "rotter," or "yellow," or a "knocker," or "conceited." The man's personal qualities loomed so large in the process of judging as to play a completely domineering role. I believe it operated in the case of these eight men as a definite inhibition to the process of "judging." It is not possible that they really "judged" his intelligence, for example. They were controlled by a predisposition, a bias, a prejudice. This predisposition was a general mental attitude toward Captain X, dominated primarily by an attitude toward him as a social associate. This attitude had been built up by countless personal reactions on the drill ground, at the mess table, in quarters at rest times and the like. And these general mental reactions were determined very generally by the overpowering effect of particular kinds of responses which he had made. I personally believe that these reactions, furthermore, were determined by the way *they* interpreted his attitudes towards them. Is it not a condition of very general prevalence that we react to another in terms of how we think he will affect us and our future. We ignore him or we pay close attention to him. We accept what he says to us or about us in terms of an attitude of confidence in how he will affect us. Our interpretation of the same identical remark made by a close friend and a hostile colleague is *determined by our general feeling of the way he probably means it.* His responses are to us symptoms of what he wants to have happen to us. I shall intrude more of this theory on the reader later on. First let us look at another illustration of what we are discussing.

In Table VIII we have another typical case, that of No. 4 rated by No. 37 and No. 38. At two different conferences No. 38 rated him "9," that is mediocre; No. 37 rated him "3" each time. No. 4, however, was rated by No. 37 as a "lowest" man in each of the 5 qualities—3, 3, 3, 3, 8, giving a total of 20, the lowest rating a man

can be given. Thus this is probably a case in which we do not have an accurate and direct comparison between two scale-men and a third man, for the "3" men on the two scales are the same. In such a case it seems clear that there must be influencing the judgment a general attitude toward No. 4 that is such as to preclude careful analysis of his separate qualities. No. 4 is rated by the group as a whole as somewhat below average—the average of the conference ratings on him is 51. He stood out as an "average" man in the psychological and alertness tests. He is a college trained man and advanced rapidly in salary in the three years preceding entrance into the service. On the whole, No. 37's rating of No. 4 can be interpreted as a case in which a general attitude mistaken certainly in some particulars, contributes to an error in judgment of an officer with respect to specific qualities. Sufficient evidence is not at hand concerning such instances for us to draw large generalizations. *The suggestion comes insistently, however, that one of the most potent influences working against accurate estimates of character is the prevalence of just such general attitudes toward our associates and subordinates.*

It is very difficult to show the influence of a rater's judgment of one set of qualities on his judgment concerning another set. The statistical data compiled in this investigation have been carefully canvassed for the determination of such possible influences; the study has led to very little mass data that are helpful. It is believed that the only way in which the human aspects of this problem can be completely analyzed is by association during a considerable time with rating officers and their subordinates. My experience with the 151 officers of this study prohibited more than a very general comment on this matter.

We have brought together the slight statistical evidence that has been found to bear upon this problem. The degree of probability can be stated that an officer who is assigned to a given scale value on one quality of the rating scale will be assigned to the same scale-value on another quality. The study of the detailed tables makes it clear that the chances are about 11 to 1 that an officer who is assigned to a given scale value on one quality of the rating scale will be assigned either to the same scale-value on another quality or to the one above it or the one below it. That is the chances are about 11 to 1 that the deviation in the second scale-value will not be greater than one interval. On the other hand, the chances vary from two to one, to one to two (with the qualities in question) that the officer will be assigned to the same identical scale-value.

The data presented so far not only invalidate single judgments of character, but they also complicate the practice of using "agreement of judgments" as a criterion of the validity of the rating scale itself. We have canvassed definite examples which have shown that identical ratings may be contributed to by very dissimilar judgments; likewise that widely divergent total ratings may be based upon comparisons with equivalent scales that must have represented close agreement in judgment; furthermore, that differences in total ratings were not paralleled by differences in scale making and the like. We are fortunate in having the direct comparisons of judgments of a trait and the objective measurement of it, in the case of intelligence. The direct evidence is conclusive of the worthlessness of a preponderance of the "ratings."

But, there is another angle to this matter of subjective estimates of character. We have shown that with a most refined technique—with one so refined that it cannot be employed in general practice—ratings are not adequate measures of character. We need still to know whether this refinement in the construction of scales and in making ratings improves the case for rating.

The answer is: It does—apparently a definite amount, but yet not enough to suggest the general use of point rating scales. Turn back and compare the average differences in the official ratings with the average differences in the experimental ratings: 10 to 20 points against 6 and 7 points. A tremendous improvement was effected in the army ratings by sending instructors out from Washington to lecture to rating officers and to teach them how to make scales. There is no doubt that the 50 to 75 per cent reduction in variability of judgment was effected largely by this mass instruction.

This has important educational implications. The marking or rating of teachers and students on a general point scale, without the aid of man-to-man comparison is closely analogous to what raters did in those spring and summer official ratings in 1918. And our evidence shows they were valueless as measures of character.

Now, the instruction and the refined technique in the experimental groups enormously improved the rating, *but it was the instruction and the fact that raters did actually make and use scales in accordance with directions, that caused the improvements. It was not the added refinement of the experimental technique that brought about the improvement.* That is shown clearly by a comparison of the *Fort* Sheridan data collected by Colonel Coss and our Camp Taylor data. Colonel Coss used the

general directions of the initial form of the scale. We used two distinct refinements at Camp Taylor: first, making an original list of at least 25 persons, second, of ranking the original list for each quality separately. I cannot see that the refined technique actually improved the results at all. The average differences, for example, are just as large in the case of Taylor data as in the case of Fort Sheridan data. The real nub of the matter is, I believe, that the errors in judging complex traits cause variations in independent judgments so great as to more than offset any reductions in variability of judgment due to improved technique.

Rating of Character Nearly a Chance Event

The examples we have studied in the past few pages reveal many of the attributes, indeed, of a chance situation. We should seriously consider, I believe, whether the making of a judgment of the character of our fellows does not closely approximate such conditions. I have considerable correlation evidence which bears directly upon that thought.

The correlation between officer's ratings and scores made upon the army psychological test were computed for 15 lots of 300 officers each, 4500 officers in all. The 15 lots were taken at random from 100,000 officers, one-third second Lieutenants, one-third first lieutenants and one-third captains. I assume that there is sufficient overlapping in the abilities under examination (ratings and performance) to lead to the expectation of a correlation of 0.5 to 0.6 between the two measures. What do we find? In each case r was less than 0.05. Most of them were 0.00. Obviously, the July official ratings were completely a matter of "chance." Apparently one might as well have numbered his men and assigned ratings by drawing balls from a bag as to rate as was done in July, 1918.

How much was the situation changed by the instruction and refined technique of the Camp Taylor experiment? The coefficients for 9 correlation tables which we tabulated for psychological and alertness test scores and ratings (number of cases varied from 35 to 137) were respectively: 0.08, 0.08, 0.09, 0.11, 0.14, 0.15, 0.20, 0.21 and 0.23; average 0.15. Hence, while we did obtain a relatively better measure of an officer's traits the difference was slight. A correlation of 0.15 implies a very wide divergence from close correspondence. It is a very "low" correlation. Several of the "experimental" correlations, indeed, were nearly pure chance situations.

How can the reliability of a rating be increased, if not by improving the technique of scale-making and rating? Clearly by getting many independent ratings on a person and averaging them. In the Camp Taylor experiment we were able to do that in an exceptional way.

Averaging the ratings on 22 officers (number of independent ratings on an officer varying from 3 to 13) and correlating with psychological test score gives for three groups:

22 officers $r = 0.48 \pm 0.07$
37 officers $r = 0.51 \pm 0.09$
126 officers $r = 0.36 \pm 0.05$

Here we have a striking example of the effect of getting many judgments and averaging them. The judgment of the individual, taken by and large is of little value. The judgment of the mass is close to the truth.

(Further evidence and a summary interpretation will appear in the February issue.)

THE RELIABILITY OF RANKINGS BY GROUP INTELLIGENCE TESTS

DENTON L. GEYER,

Chicago Normal College

When a large number of persons are ranked according to their intelligence, will one group test of intelligence place them in about the same order as another? If school children are to be assigned to classes on the basis of intelligence, will all tests place a child in the same class, or will the class-section assigned to a given pupil vary with the test used? It is the purpose of this paper to discuss some of the evidence regarding these problems which may be secured by giving two group tests to the same pupils.

The Otis Intelligence Test and the Illinois Examination were given by the same person in the junior high school grades of the Chicago Normal School during 1919 and 1920, and when 120 of the pupils were ranked on the basis of scores in the two tests—using only the intelligence division of the Illinois Examination—the median change of rank from one test to the other was found to be 18 places. The maximum change which could have been effected throughout the group was 60 places and the change left to chance would be 40 places. Six pupils changed rank more than 60 places; 37, or 30.8 per cent, less than 10 places; and 15, or 12.5 per cent, less than 5 places. The

TABLE I.—AMOUNT OF DISAGREEMENT BETWEEN TWO INTELLIGENCE TESTS IN DIVIDING ONE HUNDRED TWENTY PUPILS INTO FOUR SECTIONS ACCORDING TO ABILITY MENTAL

Order of intelligence according to Otis test	Comparative results from Illinois examination					
	Number displaced one section or more	Number beyond middle of adjacent sections	Number displaced two sections or more	Number beyond middle of second section	Number displaced three sections	Number beyond middle of third section
Section A	13	11	6	4	3	1
Section B	19	11	3	1	1	
Section C	19	11	3			
Section D	11	5	3	2		
Totals	62	38	15	7	4	1

Section A contains the thirty brightest pupils as revealed by the Otis scores, Section B, the thirty next brightest, and so on.

coefficient of correlation between the two sets of scores, by the rank difference method, was 0.642. If these 120 pupils had been divided on the basis of the intelligence scores of one test into four class-sections of ordinary size, 51.6 per cent of them would have been in the wrong section according to the other test, and 31.8 per cent of them would have been out of place by an amount equal at least to half the range of such a class-section.

The Thurstone and the Brown University tests when similarly given to 54 students of college freshmen grade showed a median change in rank of 6.7 places, as compared with a maximum possible change of 27 and a random change of 18. Twenty-four students changed rank less than 5 places, and 33 less than 10 places. Dividing these students into two classes in the order of their scores on one test would have placed 26 per cent of them in the wrong class according to the other test, and would have put 5.5 per cent of them out of place by as much as half the range of such a class. The correlation between scores is 0.74. A sophomore group of 64 students when given these two tests showed a median change in rank of 10.4 places, with 14 whose change of rank was less than 5 places, and 31 whose change of rank was less than 10 places, but with 5 whose change of rank was more than 30 places. The correlation between these scores is 0.613. Dividing the sophomore group into two classes on the basis of the scores in one test would have placed 32.8 per cent of them in the wrong class according to the other test, and would have put 6.3 per cent of them out of place by at least half the range of each class so formed.

These results may be compared with those secured by J. A. Clement in giving five of the group intelligence tests to 49 students in Northwestern University.[1] The Pearson correlations he secured were: Army-Thurstone, 0.60; Army-Otis, 0.57; Army-Pressey, 0.36; Army-Indiana, 0.36; Otis-Thurstone, 0.46; Otis-Pressey, 0.44; Otis-Indiana, 0.34; Thurstone-Pressey, 0.25; Thurstone-Indiana, 0.25; Pressey-Indiana, 0.22. It is here seen that the Indiana Mental Survey test correlates with none of the others by as much as 0.40, and that the Pressey Cross Out test has no correlations as high as 0.45. None of the other correlations can impress us as remarkably high when we remember that in each comparison we are presumably considering two measurements of the same thing.

[1] Clement, J. A.: Use of Mental Tests as a Supplementary Method of Making School Adjustment in Colleges. *Educational Administration and Supervision*, November, 1920, 6, pp. 433–444.

That they are actually measurements of the same thing is rather difficult to believe when we think of the much higher correlation found to exist between two *forms* of the same test. For example, Otis[1] reports a correlation between his Forms A and B of from 0.74 to 0.94, while Snarr,[2] using this material with 306 pupils found a correlation of 0.79, and Colvin[3] in computing the relationship in about fifty schoolrooms found a relation between these forms running as high as 0.90 and averaging 0.83. Between the two similar halves of the Brown University Test, Colvin[4] has also found a correlation of 0.76. Comparing this agreement with the wide variation cited above leads one to doubt somewhat that the different tests, though all called "general intelligence tests," are really measuring the same element in the pupil's endowment.

Significance of These Differences

The importance of this variation among the tests probably depends upon the purpose for which the results are to be used. For guiding the pupil in the choice between two studies—a foreign language and manual work—the Otis scores have proved by a two-year trial in the junior high school named above to be of real practical value. With a few adjustments in cases of extreme divergence from scholarship records, this classification has proved workable and in general satisfactory. Of course, it may be said in objection that this fact does not show that the tests picked out pupils of intelligence but rather that they selected pupils of superior "literacy"—that this case is but one more bit of evidence that the so-called intelligence tests are really language tests. It might possibly be said further that for vocational, rather than educational, guidance these tests cannot be expected to function successfully until the vocations are classified as to the familiarity with language forms which is required in each, and that even if the tests worked then, they would not thereby be proved to be reliable intelligence tests but, rather, reliable tests of literacy. Even so, there may be a relationship close enough between intelligence and

---

[1] Otis, A. S.: An Absolute Point Scale for the Group Measurement of Intelligence. *Journal of Educational Psychology*, May, 1918, 9, pp. 237–261.

[2] Snarr, O. W.: Reliability of General Intelligence Tests in Classifying High School Pupils. Unpublished master's thesis, University of Chicago, June, 1919.

[3] Colvin, S. S.: Some Recent Results Obtained from the Otis Group Intelligence scale. *Journal of Educational Research*, January, 1921, 3, pp. 1–12.

[4] Colvin: Educational Tests at Brown University. *School and Society*, 10, p. 27.

linguistic ability to allow us in many situations to consider these scores as real, even if indirect, indices of intelligence. There is no lack of evidence that persons selected by tests very similar to those under discussion were found by trial to be the persons most proficient in types of work making little use of written language or other symbols. The Army Intelligence Tests selected men in a way that corresponded very closely with the selection made on the basis of general military value by officers knowing the men well. For example,[1] in twelve companies the average correlation between rankings by intelligence tests and rankings by officers on the basis of soldier value was 0.536, and in seven of the twelve companies it ranged from 0.64 to 0.75. A great deal of evidence of this kind could be cited from the records of the army psychologists. It seems in this connection to lead toward the conclusion that, in consideration of the comparatively small amount of use which the common soldier makes of written symbols, the Army Tests were of a truth measuring some quality other than literacy which was valuable in practical life situations; and that, in consideration of the fact that the correlations would always be kept low by the number of qualities besides intelligence which make for military efficiency and of the further fact that there is no other quality which the tests from their construction could reasonably be supposed to be measuring, the Army Tests were to a large degree genuine measurements of intelligence. Now since one of the tests used in the above school experiment served as a principal basis for the Army Test,[2] it seems not improbable that it, too, measures intelligence with sufficient accuracy to be of frequent practical value, especially in situations where the discriminations demanded are not too fine.

If group intelligence scores were to be used for classifying pupils into small groups of homogeneous ability, we could apparently expect a great many mistakes, but the real significance of this would depend upon how *far* a given pupil is out of place, how serious for the purpose in hand is such a displacement and, in a practical sense, upon how much better even such a classification is than the hit or miss grouping which usually prevails. As a matter of fact, nothing is commoner in educational literature at present than favorable and even enthusiastic reports of experiments in classification on the basis of scores in some group intelligence test.[3] Disregarding the possibility that where the

[1] Yoakum and Yerkes: "Army Mental Tests," p. 30.
[2] Yoakum and Yerkes: "Army Mental Tests," p. 2.
[3] Jordan, R. H.: An Example of Classification by Group Tests, *Educational*

plan fails the experiment is not written up, this would seem to show that great accuracy in ranking the pupils is not essential, at any rate for a most noticeable improvement over present practice.

Advising students away from such abstract studies as algebra or Latin and into such concrete activities as those of the commercial course is a use to which the intelligence tests have been successfully put.[1] As noted above, the possibility that they test only that type of intelligence which works through symbols does not stand against them here, for that sort of ability is, of course, just what we want to find in this case. Deciding whether a student can carry extra courses without overworking[2] is also a use of the tests with which their alleged symbolic character will not greatly interfere.

Comparing intelligence scores with scholarship in the Normal School shows correlations as follows: in the college freshmen group, Thurstone scores with semester grades, 0.41; Brown scores with semester grades, 0.56. For 14 students the rank in scholarship differed from the rank in the Brown University Test by less than 5 places; for 32 by less than 10 places; and the median difference in rank is 8.3. For 12 students the rank in the Thurstone test differed from the rank in scholarship by less than 5 places; for 24 by less than 10 places; and the median difference in rank is 11.5 places. In the junior high school the correlations are: Illinois Intelligence scores with school marks, 0.25; Otis scores with marks, 0.32. But since school marks depend on many things besides intelligence (industry, attitude, home conditions, etc.), these low correlations can hardly be taken as seriously calling into question the validity of the test results.

Though failure to confirm teachers' marks is but an indifferent criticism of intelligence tests, failure of one test to confirm the findings

---

Administration and Supervision, April, 1920, 6, pp. 198–201.

Tildsley, J. L.: Some Possibilities Arising from the Use of Intelligence Tests. *Bulletin of High Points in the New York Schools*, June, 1921.

Miller, W. S.: General Intelligence Tests. *School Review*, February, 1920.

Branson, E. P.: An Experiment in Arranging High School Sections on the Basis of General Ability. *Journal of Educational Research*, January, 1921.

Stetson, P. C.: Homogeneous Grouping in the First Year of a Five Year High School. *School Review*, May, 1921, 29, pp. 351–365.

[1] Tildsley, J. L.: Some Possibilities Arising from the Use of Intelligence Tests. *Bulletin of High Points in the New York Schools*, June, 1921.

[2] Beeson, M. F.: Certification of Teachers by Means of Mental and Standard Educational Tests. *Educational Administration and Supervision*, November, 1920, 6, pp. 471–475.

TABLE II.—DISTRIBUTION OF PUPILS IN TWO GROUP INTELLIGENCE TESTS

A—120 Junior High School Pupils

Scores in Otis Test	Scores in Intelligence Division of Illinois Examination					
	Below 68	68–81	82–95	96–109	110+	Totals
135+	..	1	2	4	3	10
115–134	..	4	9	8	7	28
95–114	2	11	21	4	3	41
75–94	7	15	10	2	..	34
Below 75	3	3	1	..	..	7
Totals..............	12	34	43	18	13	120

B—54 Normal College Freshmen

Scores in Thurstone Test	Scores in Brown University Test					
	Below 35	35–44	45–54	55–64	65+	Totals
120+	..	..	..	2	3	5
100–119	..	..	1	4	1	6
80–99	..	2	7	8	1	18
60–79	1	8	8	3	..	20
Below 60	1	2	2	..	..	5
Totals..............	2	12	18	17	5	54

C—64 Normal College Sophomores

Scores in Thurstone Test	Scores in Brown University Test					
	40–49	50–59	60–69	70–79	80–89	Totals
120+	..	..	..	3	1	4
100–119	..	2	2	3	3	10
80–99	..	6	14	12	..	32
60–79	..	10	4	1	..	15
Below 60	2	1	..	..	..	3
Totals..............	2	19	20	19	4	64

of another is more important. The extent to which rankings by a given group test varied from rankings by another, in the data cited above, would seem to show that, as indicated in the recent symposium in this journal, there is much yet to be done before group intelligence tests can be very fully relied upon for trustworthy placing of individual pupils.

THE DEVIOUS PATH OF SLOW WORK

GRACE E. BIRD

R. I. College of Education and R. I. State College

Only recently have teachers come to realize that accuracy is not conditioned by slow work. A class experiment in adding made by Thorndike[1] and extending through several years indicates that a very close relationship exists between rapidity and accuracy. Among six-hundred seventy-one students variations were considerable, but the quickest sixty-five averaged one hundred additions per one hundred seconds. The slowest averaged only one-fourth as many. The sixty-five individuals who added the most rapidly made seven errors per thousand additions. The twenty who were slowest made an average of seventeen and one-half errors. Similar relationship is shown throughout the intermediate speed groups, and is permanently characteristic there also.

Through practice one hits upon short cuts or "kinks" as they are called by the industrial worker, thereby eliminating superfluous motions and varying factors,—hence the improvement in speed that comes through practice. According to Gilbreth,[2] however, fast motions are different in *character* from slow motions. The learner, therefore, should be encouraged to attain standard speed of motions as early as possible. If these motions are such as cannot be made by the beginner at standard speed, rapidity should approach as nearly as possible that used by the expert. Otherwise the habit may be initiated incorrectly. Also, the worker in seeking speed later may find that the different motions may cause retroactive inhibition, as in other interfering habits, not well-automatized. Jesperson, the Danish philologist found the rate for optimum initial speed in teaching languages, also to agree with these conclusions. In industrial practice the learner may be encouraged to approximate standard speed by giving him work in which the finest quality is not essential. Eventually, accuracy of method and speed occur simultaneously with good quality. In other words, if the method and the speed are taken care of, the quality will take care of itself.

By standard speed is meant not always high speed, but that rate of speed which will produce the best results efficiently. Undue haste

[1] Thorndike, E. L.: Relation Between Speed and Accuracy in Addition. *Jour. Ed. Psych.*, Vol. 5.

[2] Gilbreth, F. B. & L. M.: "Applied Motion Study," 1917, Chap. VI.

is apt to arouse such emotions as anxiety, fear, or annoyance which invariably tend to interfere with rational processes. If, however, the child's work in arithmetic or any other subject requiring both speed and accuracy be properly focalized and motivated through play stimuli so that it will seem worth his while to exercise optimum effort, he may attain speed very early in the learning process. In arithmetic, approximations of the answer rather than finding the exact result give him an opportunity to attain initial speed in the same way that the industrial beginner may approach standard speed if given work in which the finest quality of workmanship is not essential.

Gilbreth, in learning to lay bricks, observed that his teacher employed three sets of motions to do the same thing. One was the demonstrating set used for teaching, the other two were employed in his own work, one being slow and the other fast. He used different motions when working slowly than when working rapidly because of the different muscle tension involved. In the latter instance centrifugal force, inertia, momentum, combination of motions, and play for position functioned favorably. When there was no emphasis on speed he was differently affected by these variables.

In mental processes, also, there is a difference between rapid adjustment and slow adjustment. The distinction may be realized by the most casual introspection. Although adding is a familiar process it is very complex. In order to add eight and nine on paper, for example, the individual first perceives visually the number eight, at the same time perhaps experiencing one or more images involving associations depending upon his apperceptive background. This process is repeated for the number nine and for the product seventeen. Furthermore the product may be almost subconsciously resolved into other element combinations such as ten and seven, five more than a dozen, etc. The act of writing the number may attract the writer's attention to that motor performance with its own complex elements. The longer one delays the completion of the act the larger the number of "irrelevant bonds" realized. In slow addition a person may even revert to wasteful habits of childhood such as counting on the fingers, lip movement, vocalization, etc. In rapid calculation learned through properly focalized practice, such irrelevant matters are crowded out through the exercise of inhibitory processes. The first perception of the numbers set off the automatic response of the product, with the elimination of useless and wasteful intermediate performances.

Recently one hundred college students were tested by the writer

with slow and rapid adding of examples taken from the Courtis research tests. For two minutes the students were required to work as quickly as possible. The median number of errors was found to be three, the quartile deviation 0.5. The students were then asked to continue adding. This time they were cautioned to work slowly and accurately. The median number of errors was four, the quartile deviation 0.8. The workers were then requested to describe everything that entered their minds during the rapid adding. Only five individuals recorded conscious distractions of any kind. The others stated as their central thought a desire to get the answer, or to add as rapidly as they were supposed to. When required to record their thoughts as experienced during slow adding all but three mentioned distractions. These included variety of imagery, adding by combining units rather than by combining groups, consciously unnecessary repetitions of sums obtained in the process of adding a column, emotional disturbances, physical uneasiness, observation of environmental stimuli, halting uncertainties regarding the sum of certain numbers, forgetfulness of the sum already found, losing the place, slight amusement at the experiment, and fatigue.

If it were possible to draw accurate motion paths of these distractions the result would be a tangled skein as intricate as the motions of the slow industrial worker. If this vagrancy of attention occurs in individuals who have learned to add well enough to enjoy their skill, it should be even more evident in the case of the child who in the process of learning to add is only too ready to be diverted by outside stimuli from a difficult and irksome task in the stage when it is neither novel, nor yet pleasantly automatic. Continual shifts of attention to distractions might easily occasion the fatigue experienced by some of the individuals during the writer's experiment in addition. Further investigation might show decreased efficiency even more marked than the reduction of accuracy from a median of three errors to a median of four. The larger percentage of errors during slow adding and the variety of irrevelant mental content indicate that in some way the nature of the work is different from that of rapid adding.

In reading, also, if the by-paths of articulation, inner speech, eye and throat tensions, auditory, motorizing mechanisms, and imagery of the slow reader could be reproduced and compared with the direct route of the rapid reader, the relationship would no doubt parallel the comparison between slow and rapid adding.

In J. A. O'Brien's[1] experiment, photographic records were made

of the eye movements of ten pupils in grades III to VIII before and after training in silent reading. A study of the records showed that the improvement on the physiological side was effected chiefly by a lessening of the number of the fixation pauses rather than a decrease in the duration of these pauses. The development of speed was also accompanied by a marked decrease in the number of regressive movements and by the setting up of habits of regular rhythmical eye-movement. This adds evidence to the assumption that slow work is of a different character from quick work.

As has already been pointed out by M. A. Burgess,[1] scales for the comparative attainment in reading measure quality, difficulty, or amount, though reading is not easily measured by scales for quality or scales for difficulty. It is measurable by scales for amount. It is probable that difficulty will be indirectly measured eventually through a series of carefully-graded tests for amount, thereby following the law of the single variable as recognized in scientific measurement. This single variable (amount) obviously involves speed.

In a previous experiment[2] by the writer in giving standard tests to a whole school the highest correlations between tests occurred between comprehension and speed in Kansas Silent Reading and between speed and accuracy in Courtis arithmetic. In handwriting, however, a minus correlation was found between speed and legibility probably because the children had been trained to write slowly, and were therefore disturbed by the effort to inhibit superfluous motions. Rapid drill from the beginning focalizes and initiates habit with a minimum of waste.

"L'éxercice abrège le calcul, parce qu'il modifie le travail, non seulement au point de vue quantitatif, en accroissant la vitesse d'éxécution des opérations élémentaires et la vitesse de transition d'une opération à l'autre, mais et surtout au point de vue qualitatif, c'est à dire en transformant la *nature* du travail."[3]

Pupils should think in terms of results more than in terms of the process. This economical method encourages speed and is more conducive to concentration because in less danger of distraction elements which tend to alter the character of the work.

Conclusion.—Fast motions are essentially different from slow motions not only in industrial but in intellectual work.

---

[1] *Twentieth Yearbook*, Nat. Soc. for the St. of Ed., Pt. II.

[2] Bird, Grace E.: A Test of Some Standary Test. *Jour. Ed. Psych.*, Vol. II, No. 5.

[3] Foucault, M.: L'Etude Scientifique du Travail Mental Specialement Dans le Travail d'Addition. *L'Annêe Psychologique,* Tome XX, p. 125.

CONSTANCY OF THE STANFORD BINET-IQ AS SHOWN BY RETESTS

JOHN L. STENQUIST

Bureau of Reference, Research and Statistics Board of Education, New York City

In the September 1921 issue of this Journal appeared a summary of six reports[1] on the above topic including a study reported to have been made by the writer, and another by Miss Fermon. While mention was made of the fact in a footnote, it should be made clearer that the same cases are involved in both reports but the data were treated in a somewhat different way in each case. In this article the "conclusions ignore the data of Stenquist and Fermon which—must be unsound," in view of the contradictory results reported by the other workers. We are anxious to be the first to express our gratification, at the higher constancy found by other investigators. In fact, it was precisely because of the disappointingly low constancy found by us that the complete report has been withheld from publication in the hope that, other, and more encouraging ones would appear. We yield to none in our insistence upon the importance of proper standards of qualification for mental testers, but we do not feel there is necessarily final ground for admitting the unsoundness of our data. Frankly, however, *we hope they are unsound.* We fully agree that the other reports do strongly tend to cast doubt upon the validity of our results, and naturally the five reports summarized in the article referred to are therefore of particular interest to us. Our tests were given by four persons, and errors made by any of these may of course be responsible. Their training and experience was as follows:

One, a Smith College graduate and graduate student at New York University, has acted as examiner for the Public Education Association for several years, giving hundreds of Binet tests, and hence her proficiency was unquestioned.

The second examiner is a Vassar graduate where she had substantial psychological training. At least 20 Binet tests were given there by her under close supervision. Following this she had the experience of testing between 50 and 60 cases in a psychological clinic in New York City. After this she gave approximately 40 Binet tests in a survey

[1] Rugg, Harold and Colloton, Cecile: Constancy of the Stanford-Binet as Shown by Retests. *Journal of Educational Psycholgy,* September, 1921, pp. 315–322.

by the Department of Ungraded Classes in New York City. All this experience plus a thorough psychological training should make her more proficient than many examiners.

The third examiner is also a Vassar graduate, where she had had 3 years of work in various branches of psychology, and at the time the present study was conducted she was taking a graduate course in psychology at Columbia University. In connection with the course in applied psychology at Vassar College she had given about 25 Binet tests, during a period of 9 months. The first of these tests were given in the presence of the instructor and the results in the remainder were checked by the instructor, in so far as that is possible.

The fourth examiner, also a graduate student, had given at least 200 tests prior to this experiment and had had thorough college and clinical training.

Whether or not our examiners were competent can only be inferred. That our larger differences may be due to the foreign character of the population tested seems most likely, however, as in our group the language factor was a serious one. If a pupil who lives in a home where English is not spoken is tested at the beginning of school, say at age 6 to 7—and then retested after a period of 6 months to 18 months in school where the English language is acquired, it is reasonable to suppose that this knowledge of English will improve his score appreciably—as much as the improvement shown in our retests.

Thus while on the whole we too would prefer to assume that in some way the technique of our examiners differed sufficiently to explain the differences, rather than to destroy our confidence in the fairly high average constancy of the Stanford-Binet test, the language-difficulty factor alone seems adequate to explain our higher retest scores. Even with the assumption that the Stenquist-Fermon data are unsound, however, there still remain some troublesome points in the matter of the constancy of an IQ. Leaving our data entirely out of consideration for the moment we may still note the wide range—from −20 IQ to over 20 IQ in the Terman data, from −15 IQ to 17 IQ in the Rugg-Colloton data, and from −14 IQ to 15 IQ in the cases of Garrison. Does this not mean, that when we cite the case of a pupil tested within say, 6 months to 18 months, *the IQ assigned to him may be wrong by as much as* 20 *or more points?* To be sure it is chiefly a question of *how often* this will occur, but the disturbing fact is that this can and does occur at all. Even if we limit it to the large error of, say, 'not more than 15 points wrong,' it still occurs too frequently for comfort. The

percentages of cases which differed 15 points or over as shown in the article referred to are:

For Terman's data: in 29 out of 435, or in about 7 per cent of the cases.

For the Rugg-Colloton data: In 6 out of 137, or in about 4 per cent of the cases.

For Garrison's data: In 1 out of 62, or in about 2 per cent of the cases.

In our data this percentage rises to 11 per cent, which in the light of the other data seems too high. But whether it is 2 or 7 or 11 children in a hundred, in whose cases we make this huge blunder, it is serious. Assuming adequate proficiency of all testers the imperfect reliability of our scales of course also contributes to the unreliability of our constancy figures. That the Intelligence Quotient is very closely constant for each child seems doubtful in view of these wide ranges, and the relatively high reliability of Binet test, no matter what may be the case "on the average." In the Stenquist-Fermon data if we eliminate the 26 children who differed by 20 or more, the distribution is not markedly different from that of Terman's data. It is these 26 cases[1] that look the most questionable. We shall await with much interest the findings of other workers.

---

[1] Are these 26 cases those having language difficulties? This should be ascertained. H.O.R.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY CECILE COLLOTON

Department of Educational Psychology, The Lincoln School of Teachers' College

INTELLIGENCE TESTS

The Results of Repeated Mental Re-examinations of 639 *Feeble-minded Over a Period of Ten Years.* F. Kuhlmann. Journal of Applied Psychology, 1921, September, 221–224. A study of mental age growth curves and the constancy of the IQ for the four groups of the feeble-minded, each group being studied separately. Complete data given.

The Intelligence of Chinese Children in San Francisco and Vicinity. Kwok Tsuen Yeung. The Journal of Applied Psychology, 1921, September, 267–274. Results of testing 109 Chinese children with the Stanford-Binet Test. Details are given in eight tables. Comparison made with Terman's data on American children.

A Comparison of Brahman and Panchama Children in South India with Each Other and with American Children by Means of the Goddard Form Board. D. S. Herrick. Journal of Applied Psychology, 1921, September, 253–260. Racial differences in general intelligence. Comparison of the results of tests given to 355 high caste Indian Children, 355 low caste, and 1572 American children.

Pictorial Completion Test II. Wm. Healy. Journal of Applied Psychology, 1921, September, 225–239. The picture completion test as the fairest test of apperceptive abilities. Description of test; directions for giving and scoring test; and norms of performance.

A Cycle Omnibus Intelligence Test for College Students. L. L. Thurstone. Journal of Educational Research, 1921, November, 265–278. Description of the selection and cycle arrangement of six tests to be given to college freshmen. Norms of performance for the freshmen of a number of engineering and liberal arts, colleges and normal schools.

The Case for the Low IQ. J. L. Stenquist. Journal of Educational Research, 1921, November, 241–254. Criticism of the narrow, academic nature of present-day intelligence tests. Discussion of other kinds of "general" intelligence illustrated by tests of mechanical ability.

Where Test Scores and Teachers' Marks Disagree. Mary B. Lindsay and Ruth S. Gamsby. The School Review, 1921, November, 678–687. Special studies of 46 cases showing a wide difference between the score on Terman group test and the average of teachers' estimates of work of each student in each subject. Binet test used to confirm group test score. Explanation for divergence given in each case.

The Grading and Promotion of Pupils. Chas. B. Willis. Journal of Educational Method, 1921, November, 90–95. The value of mental measurement follow-up work; what has been accomplished in the Alexander Taylor School of Edmonton, Alberta, Canada.

Educational Tests

A First Report on Two Diagnostic Tests in Silent Reading for Grades II to IV. Luella C. Pressey. Elementary School Journal, 1921, November 204–211. An analysis of the silent reading problem in the lower grades followed by a description of two tests, one for speed and the other for vocabulary, to be used as diagnostic tests. Information also given to show how the tests were validated and to illustrate the practical use of the tests and the interpretation of results.

Comparative Scoring and Recording of Educational Tests. E. E. Lindsay. Educational Administration and Supervision, 1921, November, 427–432. Description of a percentage system of translating scores of standardized educational tests. Diagnostic possibilities of the system illustrated by actual cases.

The Measurement of High School English. Edward Wm. Dolch, Jr. Journal of Educational Research, 1921, November, 279–286. A defense for the amount of time given to high school English. Why the results of English teaching cannot be adequately measured.

Measuring the Efficiency of Teachers by Standardized Tests. Samuel Brooks. Journal of Educational Research, 1921, November, 255–264. Rating the teacher according to progress made by pupils as measured by standardized tests. Illustrations of the practical working of the plan.

Tests for Special Abilities

The Construction of Tests for Discovery of Vocational Fitness. Frank Watts. Journal of Applied Psychology, 1921, September, 240–252. A classification and discussion of the tests already in use. Guiding principles in the construction of such tests.

Methods for the Selection of Comptometer Operators and Stenographers. M. A. Bills. Journal of Applied Psychology, 1921, September, 275–283. Report of a study made with certain tests of the Bureau of Personnel. Research of Carnegie Institute of Technology, to determine whether the tests would (1) eliminate failures, and (2) select sure successes. Satisfactory results given in detail.

Miscellaneous

Three Refinements of Method in School Surveys. Florentino Cayco and Sidney L. Pressey. Educational Administration and Supervision, 1921, November, 433–438. Report of a survey of Grades 1, 2, and 3 in three ward schools. Educational efficiency shown best by "ability grade table," evenness of development in all subjects, and correlation between ability and achievement in individual cases.

The Relative Standing of Mathematical and Non-mathematical Pupils. John A. Marsh. Educational Administration and Supervision, 1921, November, 458–466. Results of a study of 115 pupils in the Boy's English High School, Boston. Two groups—one studying no mathematics, the other studying mathematics in the first year. Groups almost exactly the same in first year work. Mathematical group decidedly superior in work of second and third years.

Mind-set and Learning. William H. Kilpatrick. Journal of Educational Method, 1921, November, 95–102. Part I of a popular presentation of the laws of learning.

Filmed Geometry. Charles H. Sampson. Journal of Educational Method, November, 1921, 116–117. The place of the educational film in the class-room and especially in evening schools.

Mental Types, Truancy, and Delinquency. Edgar A. Doll. School and Society, 1921, November 26, 482–485. Truancy and consequent delinquency in large part the fault of the public school system. Need for a scientific classification of children according to individual differences in mental type, and differentiated courses of study.

Investigations Undertaken by the Society for Experimental Pedagogy in Denmark. Christian Hansen Tybjerg. Journal of Educational Research, 1921, November, 301–307. Brief mention of a number of investigations, physical and psychological, conducted by the Society for Experimental Pedagogy, with the results of each.

Minor Studies in Educational Psychology. Francis Gaw. Journal of Applied Psychology, 1921, September, 284–286.

1. *School Ratings and Moving Pictures.* Influence of movies on the conduct and school ratings of 337 children in a suburb of Boston. Practically no relation.

2. *Relation of Stanford Tests and Dearborn Maze Tests.* Correlations between Stanford-Binet Scores and Dearborn scores of 77 patients at the Boston Psychopathic Hospital grouped according to diagnosis. Comparison of Dearborn scores of 77 patients and 36 normal adults connected with the hospital.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

1. *A new and important text in General Psychology* of particular interest to educational psychologists is the new text by Woodworth.[1]

The usual text in psychology has seemed to many psychologists working in the field of education to offer comparatively little which could be applied to the solution of educational problems. There were, of course, two interpretations of this fact; one, that general psychology was by nature not susceptible of direct application and the other, that the type of psychology ordinarily represented in general texts and courses was not of the character which could readily be applied. Woodworth's text demonstrates that the second explanation is more nearly the correct one. It represents distinctly a type of treatment which, without much direct discussion of educational problems, illumines the processes which are involved in learning and in teaching. This will be clear from a description of the book.

The general plan of the book is as follows: After defining and delimiting the subject in a simple and clear fashion, the author opens with a discussion of reactions. He begins with the simplest reactions,——the reflexes,——and proceeds to a discussion of the more complex ones, endeavoring to avoid a break in the continuity of the discussion. The different levels of reactions are discussed with particular reference to the organization of the nervous system. The nervous system, however, is treated not as a separate topic but simply as a link in the chain of the explanation of reactions. Neurological explanations, moreover, are included at any place in the book where they are called for. In this way the whole treatment is permeated by a reference to the nervous basis of mental life.

Transition from the simpler reactions to the higher and more complex ones is made through the development of the concept of tendencies. These are the relatively permanent dispositions of the organism which bring about what are sometimes called indirect responses. They

---

[1] Woodworth, Robert S.: "Psychology, A Study of Mental Life," New York, Henry Holt & Company, 1921, p. 580.

account for such features of mental life as motives, without abandoning the fundamental notions of reaction.

The transition to a descriptive account of the various types of native reactions is furnished by a discussion of the relation between native and acquired responses. The native reactions themselves are classified under the heads of *instinct*, *emotion*, *feelings*, *sensations*, *attention* and *intelligence*. It will be seen that the order is, in a number of instances, the reverse of the usual one; thus *emotion* comes before *feeling* and *sensation* and *attention* after a prolonged discussion of activities in which they are involved. This accords wih the general mode of treatment by which the total reaction is first described as a whole and then analysed into its elementary processes. Other illustrations of a similar reversal of the usual order are found in the placing of *perception* after *learning*, *association* after *memory*, and *imagination* after *reasoning*. The same reason holds here as in the previous case.

The chapter on *Intelligence* closes the treatment of native responses and forms the transition to the description of acquired responses. As the first phase of the discussion opens with native reactions in the form of *reflexes*, so the second phase opens with acquired reactions in the form of *learning*. This is followed by *memory* including the account of the process of memorizing in some detail, of association, perception, reasoning, imagination, will and personality.

This plan reveals the general character of the book. It is behaviouristic, with a small "b." Mental life is conceived as a form of activity organically related to bodily activity, and not as a passive spectator on the scene of life. The author refuses to follow the extremists of the behaviouristic school, however, but makes reasonable use of the method of introspection and ascribes due importance to the sensory, perceptual and imaginational processes. These processes, however, are not independent elements but are functional parts of reactions. The inclusion of tendencies saves the discussion from an undue emphasis upon the simple animal-like type of reactions.

The content of the book is comprehensive. It includes the somewhat novel topics of *learning*, *memorizing* and *intelligence*, besides the usual ones. These, of course, fit very naturally into the general plan of the book and constitute part of the reason why it will prove particularly useful to educational psychology.

The whole discussion as well as the general plan is thoroughly matured, well-organized, systematic and consistent. There is nothing improvised about the book. The author uses the results of scientific

studies ranging over the whole field but presents them in a thoroughly assimilated form. The student is given the conclusions from these scientific studies without being confused with detailed debates on matters of theory. The chief issues, however, are presented in clear and simple fashion.

The style of the book is simple, direct and, in places, colloquial. This will perhaps be an attractive feature to the undergraduate student. The concessions are sometimes considerable, as in the sentences "There are lots of nerve cells," "Not that Freud would OK our account of dreams up to this point." In places the style becomes picturesque, reminding one of James: "Man is by all odds the most pottering, hem-and-hawing of animals." It seems likely that the book will in some measure, at least, counteract the tradition that the study of psychology is a very formidable and abstract affair.

It will be gratifying to many psychologists to find the author, while giving due credit to the contributions made by his research, refusing to accept the extravagances of Freud's theory. His sane and comprehensive statement of the limits of his theory should have large influence. This is but an instance of the balance and sanity of the entire book.

FRANK N. FREEMAN.

---

2. *The Second Volume on the Virginia Survey.*—Part II of this survey report[1] deals with educational tests. The purpose and scope of the measurement program are outlined in the opening pages. Local conditions necessitated a careful adaptation of test materials and standards if the survey was to accomplish its twofold purpose: (1) To present such evidence of the status of the schools as might lead to necessary action for improvement by constituted authorities. (2) To disseminate information, stimulate interest and develop understanding of the best educational methods to make for a permanent local force for the improvement of education.

The difficulties of administering the state-wide testing movement were surmounted by effective organization under the leadership of Dr. Haggerty and by the assistance of the General Education Board.

[1] Hart, Harris, President of the Virginia Education Commission and Inglis, Alexander J., Director of The Virginia Survey Staff: "Virginia Public Schools: A Survey of a Southern State Public School System." "Part II—Educational Tests." "Educational Survey Series." Yonkers: World Book Company, 1921, pp. 235.

About 16,000 different children were examined with from six to forty tests each. About 5,000 were in grades III to VII of rural white schools. Another thousand were in grades I and II of the same schools. About 6,000 white children were in the seven grades of city schools, and in the first year of high school. About 3,000 colored children were examined. Great care was exercised in the selection of schools to be tested, and in the selection and training of prospective examiners. The scoring was done by specially trained and supervised advanced and graduate students and carefully checked by the survey staff.

While Dr. Haggerty is responsible for the general plan of the reports, other members of the survey staff contributed chapters. Chapter II contains a concise preliminary statement of conclusions and recommendations which grow out of the statistical evidence submitted in the following chapters. In addition to the tabulations, graphical representations, and the other matter usually found in such reports, there is a long chapter on the criteria for evaluating tests as a basis for grouping elementary school pupils. This chapter is designed for the critical reader and gives the statistical basis of statements made in the following chapter. While most of the conclusions are of local interest, the data assembled in the volume are worth careful study and is a valuable addition to survey literature. Students of Education in other southern states will find in this volume suggestions for the solution of their problems.

L. Z.

---

3. *Light on Some Aspects of Education in England.*—Students of comparative education will find in this addition[1] to the "Modern Educator's Library" a brief exposition of the chief features, principles and ideals in English education as exemplified in the organization and curricula of schools. The material will be much more readable to those who have acquired in some previous experience, the English connotation of such terms as "Public School," "Elementary Education," not to mention "vulgar fractions,"and such grouped modifiers as "*Ordinary Public Elementary School.*" A glossary of English educational terms with American equivalents would save much

---

[1] Sleight, W. G.: "The Organization and Curricula of Schools." "The Modern Educators' Library." New York: Longmans, Green and Co.; London: Edward Arnold, 1920, p. 264.

descriptive matter and help the American reader to sense the situation described.

A brief historical introduction is followed by two chapters on organization of schools and one on buildings and equipment. Chapter VI deals with principles underlying the curriculum and is followed by four chapters dealing with particular aspects of curricula. One chapter is given to the discussion of a flexible curriculum, the feasibility of which would be enhanced by the general adoption of a "minimum curriculum of fundamentals." A chapter is given over to the presentation, analysis and criticism of "time tables" or class programs. Some of the evaluations show that differences between English and American standards lead to widely divergent conclusions with reference to the same data.

Some of the tabulations are not headed, labelled or interpreted, and the only indication of what they represent must be sought in the adjoining pages. There is a chapter on teacher training, classification and other administrative problems, only part of which is factual. Chapter XI discusses the psychological foundations of school government at some length. The next chapter is given over to brief descriptions of the status of education in other lands. This is followed by a discussion of the implications of the "Education Act of 1918." The book contains a classified bibliography of pertinent educational literature.

L. Z.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Vol. XIII February, 1922 No. 2

A STUDY OF HIGH SCHOOL SPELLING MATERIAL

JOHN A. LESTER

The Hill School, Pottstown, Pa.

Among the literature of spelling, now of considerable extent, there have been many valuable investigations made with the object of learning and describing the source of the ability to spell, its measurement, hygiene, relation with other abilities, the source of memories involved and their relative efficiency. There have been very few attempts to discover the limits of the problem, to define its exact nature, and to seek specific remedies. The waste of time and money in the teaching of spelling is a commonplace among educators. It has been variously calculated that from one to three years is lost in the school life of every child in the process of learning to spell, and the national loss has been estimated in hundreds of millions of dollars. While these estimates are more or less guesswork, the practical school problem is the double one, first of gaining efficiency, and second of saving time.

I. Problem

The investigation, some of the results of which are set forth in the following pages, was begun some years ago in the belief that a great economy of time might be effected if for a given age the extent and nature of misspellings were determined. It sought, therefore, to learn not only what words are misspelled, but also how they are misspelled; and having arrived at this knowledge, it aimed at devising by experiment and practice efficient and time-saving methods of teaching this material. The specific objects of the investigation may be stated as follows:

1. To determine what words are most frequently misspelled by the graduates of high schools and preparatory schools.

2. To determine how these words are misspelled.

3. To determine how these words may be taught and learned with a minimum expenditure of time and energy on the part of teacher and student.

II. Material

The material used was the compositions written upon subjects from their own experience by candidates for the College Entrance Examination Board's papers in English in the years 1913–1919 inclusive. This material seems to afford a safe basis for investigation for it is broad in its nature, offering free scope for free composition on matters within the experience and knowledge of the writer; it is the work of 2414 students widely scattered, resident in forty-six different states; it is the work of pupils of seventeen to eighteen years of age from every social class and every kind of school training; it represents the product of this training when it has presumably reached a uniform level—the level set by the College Entrance Examination Board. That the conditions under which the compositions were written were not found disturbing or distracting is evidenced by the fact that nearly forty per cent of the books were finished and handed in before the expiration of the time allotted.

In approximately 1,378,000 words of free composition written by 2414 different candidates 2602 words were misspelled; and these 2602 words gave rise to 14,002 misspellings. It should be observed that in this count was not included the work of candidates obviously illiterate, and of foreigners unacquainted with the English language, and that in the composition work of every individual a word misspelled was counted only once, unless the writer varied in his misspelling, in which case each variation was counted. A word obviously containing two or more misspellings (*e.g.*, *resieve* for *receive*) appears in the count as a single misspelling. The distribution of these 2602 words in relation to the 14,002 misspellings is given in Diagram No. I.

From this diagram it is seen that 10 words were responsible for more than 6 per cent of the total misspellings; 50 words were responsible for nearly 20 per cent; 100 for more than 30 per cent; 200 for 44 per cent; 300 for more than 50 per cent; and 775 for nearly 75 per cent. The plotted curve in Diagram No. II shows the distributions of the misspellings of these 2602 words.

In what follows attention is centered upon these 775 words of greatest frequency. They include all the words which were misspelled,

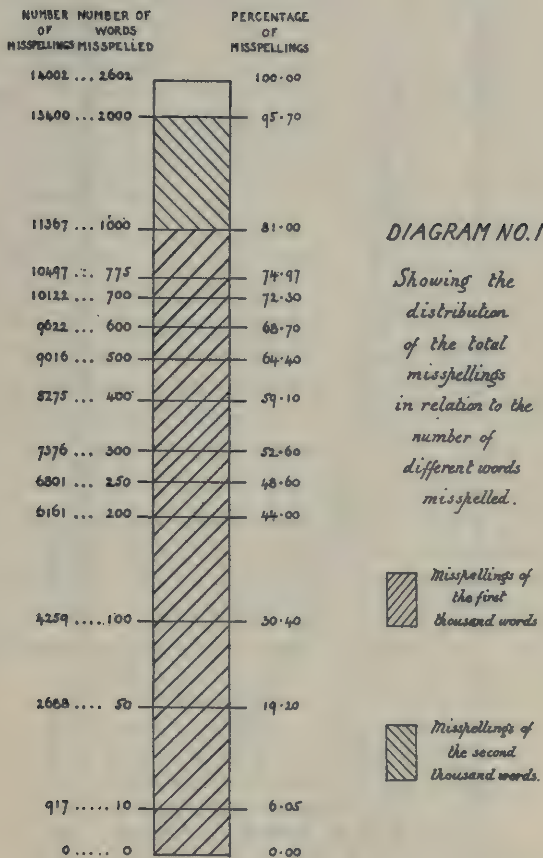
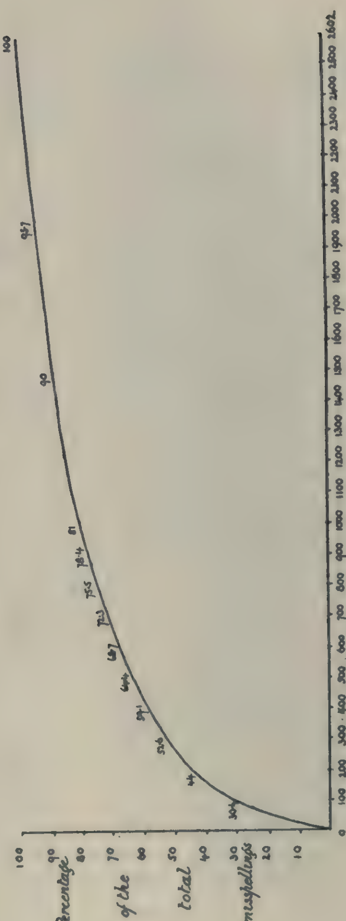

NUMBER OF MISSPELLINGS
NUMBER OF WORDS MISSPELLED
PERCENTAGE OF MISSPELLINGS
14002 ... 2602
13400 ... 2000
11367 ... 1000
10497 ... 775
10122 ... 700
9622 ... 600
9016 ... 500
8275 ... 400
7376 ... 300
6801 ... 250
6161 ... 200
4259 ... 100
2688 ... 50
917 ... 10
0 ... 0
100·00
95·70
81·00
74·97
72·30
68·70
64·40
59·10
52·60
48·60
44·00
30·40
19·20
6·05
0·00
DIAGRAM NO. 1
Showing the distribution of the total misspellings in relation to the number of different words misspelled.
Misspellings of the first thousand words
Misspellings of the second thousand words.

DIAGRAM II Distribution of the misspellings of the 2602 words misspelled.



a total of more than four times in the composition writing of 2414 candidates during the years 1913–1919. As will be seen from Diagram No. I, these 775 words occasioned 10,497 misspellings.

To attack intelligently the problem in hand it is clearly necessary to know not only what words are the causes of the trouble, but also how they cause it. It was thought advisable, therefore, from the first to record not only the word misspelled, but the form of the misspelling. When these forms of misspelling were collected, classified and studied, the existence of certain current types of misspelling in many words became evident. Many words displayed not merely one misspelling, but two, and sometimes three types of error. If the relative frequency of types of misspelling was to be determined it was necessary first to classify the errors occurring in the recorded forms, each under its proper type. When, therefore, two different errors occurred in a single word, each error was recorded with the type to which it belonged. Thus *existence* is frequently misspelled *existance*, and less frequently *exsistence*. But sometimes a student writes *exsistance*, thus making two separate and distinct errors in the spelling of the word. Again the form *comite* shows three errors; *m* for *mm*, *t* for *tt*, *e* for *ee*. In all such cases, each error for the purpose of type study was catalogued separately according to class. The 10,497 misspellings of 775 words resolved themselves into 10,853 type misspellings. Of the 775 words studied the first 50 in order of frequency of misspelling are set forth in the following list. In the first column occurs the word, in the second the number of times it was misspelled, in the third the different types of misspelling with their relative frequencies.

too	167	to 158, two 8, toe 1
its	160	it's 154, itt's 3, hit's 1, it'z 1, its's 1
believe	77	beleive 72, bilieve 3, beleeve 2, beleave 1
together	73	to-gether 51, to gether 18, togeather 4
their	66	there 37, thier 26, they're 3
principal	63	principle 59, prinsipal 2, principall 2
committee	62	comittee 22, committe 18, commitee 16, committy 10, committey 2
therefore	61	therefor 49, there fore 5, therfore 2, theirfore 5
separate	61	seperate 57, saperate 4
pleasant	59	pleasent 50, plesant 13, plessant 2
receive	56	recieve 50, recive 4, reseive 2, resceive 1
benefit	55	benifit 37, benefet 16, benefeit 5, benefite 2
occurred	54	ocured 42, ocurred 8, accured 2, occered 2
there	53	their 51, ther 2
o'clock	49	oclock 41, o-clock 8

don't	47	dont 32, do'nt 11, doe'nt 4
immediately	47	immediatly 24, imediately 7, immeadiately 7, immidiately 1
affect	47	effect 38, afect 9
business	45	buisness 32, bussiness 8, busyness 2, busines 2, bizziness 1
equipped	45	equiped 30, equipt 13, equipted 2
acquaintance	45	aquaintance 27, acquaintence 17, adquaintance 1, ackwaintance 1
discipline	45	disciplin 16, dicipline 6, disipline 4, disapline 2, dissipline 1
independent	44	independant 41, indipendent 3, indeppendent 1, independente 1
ammunition	44	amunition 42, amunishon 2, ammunittion 1
referring	43	refering 38, refferring 8, refearing 1
necessary	42	neccessary 24, nescessary 11, necesary 4, necessery 3
until	42	untill 39, un-til 3
existence	41	existance 36, exsistence 12, existense 2
principle	41	principal 39, prinsiple 2
appearance	40	appearence 33, apperance 9, appearrance 3
led	40	lead 40
across	40	accross 37, acros 3
dependent (adj.)	39	dependant 37, dipendent 1, dependente 1
extension	39	extention 37, exstension 2
occasionally	39	occassionally 26, occaisionally 6, occasionaly 4, occasonally 2, occasionly 1
surprise	38	suprise 31, supprise 7, surprize 3, saprise 2
government	38	goverment 22, govenment 12, govonment 4
acknowledge	38	acknowlege 15, acknoledge 21, acknowledg 2
lose	37	loose 35, loos 1 loze 1
effect	36	affect 34, efect 2
choose	36	chose 32, chooze 4
successful	35	sucessful 28, succesful 6, successfull 4
athletic	35	atheletic 32, athlectic 2, athelletic 1
possession	35	possesion 21, posession 10, possetion 4
opportunity	35	oportunity 28, oppertunity 8, apportunity 3
quarter	34	quater 32, quartter 2
before	34	befor 29, be-fore 5
beginning	33	begining 27, beggining 9
aeroplane	33	areoplane 18, airoplane 14, earoplane 1
sense	33	sence 31, scense 6

III. Classification of This Material

The object of the study of the misspellings of these 775 words was to evolve a classification and presentation of this material for the purpose of teaching it with thoroughness but with the greatest economy of time and energy on the part of teacher and student.

(*A*) *Nature of the Words Misspelled*

An examination of the 775 words most frequently misspelled by boys and girls of seventeen to eighteen shows that more than 50 per cent of the misspellings are of words in the grade vocabularies, and 31 per cent are of words in Ayres' list of the thousand commonest English words. And yet the joint vocabularies of these students is large and varied. Only 45 out of the 775 words occur in the 600 words which, according to Ayres, constitute more than seven-eighths of the ordinary words of written expression; 606 or 78 per cent of the 775 words do not occur at all in Ayres' thousand commonest words; and 37.6 per cent are not to be found in the combined vocabularies of 13 adults as investigated by Cook and O'Shea. About one-third of the misspellings collected are of words of three syllables; one-third misspellings of dissyllables; and one-third misspellings of monosyllables or of words of four syllables or more.

(*B*) *Nature of the Misspellings*

1. *Derivatives.*—It is important for the pedagogy of spelling to know whether the word lists presented should include derivatives, or whether they should be drawn up, as is usually the case, upon a dictionary basis. Table I shows the relative frequency of misspellings of derivatives.

TABLE I.—MISSPELLINGS OF DERIVATIVES

	Number of words whose misspellings are more than half of them misspellings due to derivatives	Number of words which show any misspellings due to derivatives	Total number of misspellings due to derivatives
Number..........	180	213	2709
Percentage........	23.2	27.5	25.0

About one-quarter of the misspellings of high school and preparatory school graduates are misspellings of derivatives.

2. *Lapses.*—There is an obvious distinction between a misspelling which is clearly due to ignorance of the correct form of the word, and a lapse or error of inattention. An effort was made to determine what proportion of the total 10,497 misspellings fall into certain species of lapses as defined by those who have studied them. The classifica-

tion adopted is not a complete one, nor are the various classes mutually exclusive; but each is fairly definite. No misspelling has been counted in more than one class, though arbitrary judgment was necessary in many cases in making the decision as to which class a particular error should fall. The classification is adopted from the work of Bawden, Hollingsworth, Winford and others. The results of this classification are shown in Table II.

The writer is of the opinion that a great many misspellings which have been placed in one or other of these 11 classes are not real lapses at all, and that the analysis of lapses is at present not sufficiently exact to make such a classification as that attempted in the table of much value. The first three categories are obviously too wide in their implications. And yet with due allowance for the fact that an exact differentiation of lapses and errors of ignorance is not possible, the table does give evidence of the great amount of misspelling which is due to carelessness and inattention. If the classification had been extended to include errors due to the influence of auditory imagery, errors due to "internal speech" or "mental pronunciation," ellipses due to previous pronunciation of the same sound, and other lapse errors defined by Bawden, the number of misspellings which might be claimed as due to lapses would be a formidable majority of the whole.

3. *Critical Point.*—It is necessary, if the teaching of spelling is to be made efficient, to know not only what words are difficult but in what way they are difficult. Accurate knowledge of the common or "popular" misspelling in words usually misspelled is one of the clear paths to the saving of time and labor in the teaching of spelling. It is of the utmost importance for us to know what proportion of the aggregate misspellings of a students of a particular age is due to a single crux in the words he frequently misspells. The modern theory of the teaching of spelling rightly places great stress upon the anticipation and prevention of error and upon making the first impression clear and strong. But unless we know in what part of the word lurks the particular error we wish to anticipate and prevent, and unless we know what emphatic first impression we wish to give, how can this teaching be effective? Teachers have generally worked from purely supposititious foundations, as if this knowledge were intuitive; and have in many instances overlooked the real pitfalls in the words they present. It is clear that in anticipating and preventing error by emphatic first impression or in removing it when it has begun to grow, the first step is to be certain what that error is.

Table II.—Number and Percentages of the 10,497 Misspellings which are Attributable to Certain Species of "Lapses"

	1	2	3	4	5	6	7	8	9	10	11
	The writing of one word for another (Gregory)	Substitution of a letter which has phonetic elements common with the required letter	Errors showing tendency to form familiar words or word parts	Errors resulting from the transposition of two adjacent letters	Errors due to the perseverance of an element already written	Errors due to the tendency to omit letters which require similar motor response	Errors due to the doubling of the wrong letter in a word containing a doubled letter	Errors due to the tendency to substitute a letter with similar visual elements	Errors due to the sense of having written an element	Errors due to anticipation of an element	Errors due to the omission of a last letter when the next has the same or a similar sound
Number........	1396	1287	1133	468	441	379	342	227	190	177	39
Percentage......	13.3	12.2	10.8	4.5	4.2	3.6	3.3	2.2	1.8	1.7	0.4

Totals 6079 misspellings, or 57.9 per cent of the total 10,497 misspellings.

Table III gives the number and proportion of misspellings which may be traced to a popular error in each word.

TABLE III

Total number of type misspellings of 775 words	Number of these misspellings which show for each word a single popular type	Percentage
10,853	8312	76.6

Thus out of a total of 10,853 type misspellings of 775 words, 8312 or 76.6 per cent are due to a single false form in the case of each word. In other words, in writing the form of these 775 words, the student is inclined toward a particular error. If that error can be forestalled, anticipated, or corrected, 76.6 per cent of the misspellings will disappear.

(*Concluded in March Issue*)

DATA ON THE TRUE-FALSE TEST AS A DEVICE FOR COLLEGE EXAMINATION

F. B. KNIGHT

State University of Iowa

Many competent instructors share the belief that the so-called "true-false" test now largely confined to mental alertness tests could be adapted to college examinations in specific subjects. If this adaptation could be made, a distinct gain in the technique of college instruction would be accomplished, for a clear saving of 10 to 90 per cent of the time now spent by the instructor upon the examination of his pupils would then be possible.

As far as the usefulness of the true-false test is concerned the crux of the problem pertains to reliability. *Does the true-false test really test?* Before drawing conclusions about the reliability of this type of examination, much more data must be available and more experimentation must be done. Gates published in this journal (June, 1921) data and excellent treatment of the data upon the reliability of the true-false tests as used in classes in Educational Psychology in Teachers' College, Columbia University. We submit further data based on the use of true-false tests in a college course in Physics.

In the winter of 1921, the Elementary Physics classes of the State University of Iowa were given a true-false test. The 182 cases ranged from freshmen in the University to seniors and graduate students.

Those were tested who had a class standing from A down to Condition.[1] Those who had failed the semester course were dropped at the end of the semester, and the test was not given until the week following. It was pre-supposed that those in the class had had High School Physics.

The test consisted of three types of problems. Some were statements which the students were to mark T, f, or U, depending upon whether they thought the statements were true, false or if they were uncertain about them. There were forty-three statements of this type, fifteen of which were based on context already studied by the class within the few months preceding the experiment and the

[1] *A* represents upper 5 per cent of work; *B* represents excellent work; *C* represents fair work; *D* represents passing work; *E* represents poor work; *F* represents failure; *Cond.* represents conditional work.

remainder, twenty-eight, were taken from the part of the class text dealing with light, which the class had not yet covered. Six were of the form of multiple response propositions, and the remaining fifteen were dissected equations, which latter division had been covered by the classes.

The following are examples of the test as given:

If statement is true put a *T* in column headed T.
If statement is false put an *f* in column headed f.
If you are uncertain put a *U* in column headed U.

	T	f	U
1. Light must fall upon the objects themselves if we are to see them.			
2. In order that a body may be seen light must pass from it to the eye, and usually this takes place along straight lines.			
3. No appreciable time is required for light to pass from one point to another.			
4. The velocity of light in water is less than in air.			
5. Electric waves have the same velocity as light.			
6. The principal focus of a mirror is that point where all rays parallel to the axis meet after reflection.			
7. It is only necessary to trace two rays from any point in the object to find by their intersection the position of the corresponding point of the image.			
8. It is possible by a continuous self-adjusting process for heat to be transferred from a colder to a hotter body.			

Of the several choices offered, check the truest one.

The reaction is
- greater than the action.
- inversely proportional to the action.
- less than the action.
- proportional to the action.

The velocity of a body at the foot of a frictionless inclined plane depends only upon
- the height of the plane.
- the slope of the plane.
- the length of the plane.

The pitch of a tone depends upon the
- amplitude of the vibrations.
- particular manner of vibration.
- quality of the vibrating body.

Below are equations with important factors omitted.
Fill in each equation so that it will represent the fact.

1. $a = \dfrac{v}{t}$ when a is acceleration.
2. $r = l$ where r is the length of the right arm of the balance.
3. $h = l$ in an inclined plane if the supporting force P acts through the length l, while the weight W is only raised against the earth's attraction through a distance h.
4. $n = \dfrac{\quad}{2}$ if n is the frequency of the cord or the number of vibrations per second.

The only novel feature in the procedure of giving the test was that the examiner read each statement, commanded it to be marked and proceeded to read the next statement. This fostered immediate decision and discouraged any proclivity to borrow from one's neighbors.

The papers were then graded according to the percental error method of computation. (See Table VI, Appendix, Rugg's "Statistical Methods Applied to Education.") The sum of scores made by each individual on questions answered is his grade. After getting the grade of each one, an individual's standing in the group is easily determined.

The test was based upon two parts of the text; firstly, that covered by the students at the time of the experiment, and secondly, that part not yet studied but containing material the pupils' knowledge of which was due to high school study or incidental learning. We have then the relative ratings of students in three ways:

1. How each student stood in comparison with his class in an examination based on physics not studied in class.

2. How each student stood in comparison with his class in an examination based on physics not studied in class.

3. How each student stood in comparison with his class in an examination based on physics gained partly from commonly shared class study and partly from physics knowledge gained outside of class.

These three ratings were correlated (Pearson Product Moment) with the marks given by the Physics instructor for the term's work, which marks were, of course, entirely independent of the experiment. The semester marks were derived by the following procedure: The mean mark of three examinations, the mean of frequent short quizzes, and a mark for the laboratory work, all expressed in percentages were added together and divided by three. The resulting figure was the term work.

The correlation between the semester grades and that part of the true-false test based on material studied in class by all students was $+ 0.455 \pm 0.07$.

The correlation between semester grades and true-false test, studied and non-studied material, was $+ 0.394 \pm 0.07$.

The correlation between semester grades and true-false test, non-studied material, was $+ 0.107 \pm 0.08$.

The size of these correlations leads one to think that the true-false test based on subject matter studied in class is not related to grade marks because of general information or general intelligence factors in the way a mental alertness test would so correlate. The true-false test based on subject matter studied is peculiarly a test of that subject for if it were not, the correlations of true-false tests otherwise constructed should correlate as highly. They do not. It is certain then that there is a genuine relationship between Physics as studied in class and scores in a true-false test based on that material.

The further question is: *Does the true-false technique admit of a close enough relationship between work done in class and scores in the test to be useful?* We do not know how high the correlation between semester grade based on many factors and examination grade based on knowledge only *should* correlate.

We can only then compare the correlation between total grade and examination, old method, with the correlation between total grade and examination, true-false method.

Gates has shown that the true-false method in Educational Psychology courses stands up as well as does the examination grades gained from compositional examinations.

Our Physics data show a correlation of $+ 0.455 \pm 0.07$ between semester grade and true-false tests of knowledge. The correlation between semester grade and the final examination, old method, was in classes reported $+ 0.92$ error negligible. It is evident that this $+ 0.92$ correlation is enlarged because one-third of the semester grade is the examination mark itself which we are correlating with the examination mark. When the examination mark is taken out of the semester grade (and this seems fair because the examination is a test of Physics work—a composite of ground covered in the quizzes and of laboratory work) the correlation between the examination grade and semester work otherwise graded is $+ 0.645$.

The true-false test correlates with semester work	$+ 0.455$
The written examination correlates with semester work	$+ 0.645$
The written examination correlates with true-false test	$+ 0.576$

When it is remembered that the section of the true-false test comprising work covered took only about eleven minutes, it seems fair to conclude that a more thorough true-false test including ingenious statements concerning laboratory technique can be expected to do as well, if not better, than written examinations with the sound advantage clearly on its side of saving the instructor's time.

This we hope to demonstrate shortly. The above data are reported further to substantiate Gates' faith in the usefulness of the true-false technique in college testing. Obviously it would be helpful for other experimenters to work with this method and report their findings.

Note.—Apart from the body of the article, I wish to call attention to an aspect of the data of statistical interest. In a true-false test on material about which the examined knew nothing, we would expect as many right answers as wrong answers by pure chance. Not every statement would be checked 50 per cent right and 50 per cent wrong, but the per cent of right and wrong answers would arrange themselves in a normal probability curve with 50 per cent right as the central tendency if enough trials were made. Then to get half the answers right on a true-false test would represent not knowledge of the subject, but the most probable operation of chance. Should not then the scale values of answers be built upon a table not from 100 per cent failing to 0 per cent failing but from 50 per cent failing to 0 per cent failing, as 50 per cent failing and more could be accounted for by chance?

The percentile error of true-false markings by 142 students on 43 statements was distributed as follows:

On matter not studied:

Percentile Error	Number of Statements
0–10	0
10–20	4
20–30	6
30–40	7
40–50	2
50–60	1
60–70	1
70–80	3
80–90	4
90–100	0

On matter studied in class:

Percentile Error	Number of Statements
0–10	0
10–20	3
20–30	2
30–40	2
40–50	1
50–60	2
60–70	2
70–80	1
80–90	2
90–100	0

As far as these data go, it seems that the fact that pure chance would warrant us to expect at least few percentile errors over 50 per cent need not concern us. The pupils know something about the test even if their knowledge makes for wrong rather than right responses. Score values made upon a table of 0 to 100 per cent error seems better than scores on a table of 0 to 50 per cent wrong.

IS THE RATING OF HUMAN CHARACTER PRACTICABLE?

(*Continued from January*)

HAROLD RUGG

The Lincoln School of Teachers College

We have discussed one good measure of the validity of single ratings of character: The correspondence of ratings of "intelligence" with objective measures of it. In the army investigation there was another, a practical criterion of promise of success; namely, "appointment to a captaincy from civil life without previous military training or experience." Men in training camps were generally given officers' commissions of second lieutenant grade; less frequently were they made first lieutenants, very infrequently captains. Rarely indeed were they made majors. A careful canvass of thousands of officers' qualification records showed that most of the men appointed at once to captaincies or majorities had had military school training, or national guard training and experience. Three thousand men who had never had any such training or experience had been made captaint from training camps. These we found in our search of the records. I assumed, therefore, that these men embodied in outstanding fashios the qualities demanded for success as officers in the army. Hence in appeared that such an appointment itself was an objective and most practical measure of the traits of the men. If the ratings are valid, our assumption runs, they should be definitely higher than the ratings of men who became captains after being second and first lieutenants.

I tabulated the ratings of 3000 "civil life" captains and of 6000 captains who had been lieutenants. The medians and first and third quartiles are given in Table XII. What do they show? They show clearly that the official ratings were not adequate measures of these officers' traits. The medians, Q_1s and Q_3s (irrespective of staff corp) for both groups of captains are practically identical. But with differences in median rating in the experimental group of 8.4 and 9.4 points respectively we find confirmation of our earlier comment that the careful construction and use of scales and the improvement in technique had made the "experimental" ratings more valid measures. As to how much more valid they were we are left in the dark. No attempt was made to treat the data by more refined statistical methods for it was felt that the measures and the treatment described in the

foregoing pages had thrown more light on the matter than these data could.

In the course of the investigation several hundred correlations were computed, for groups varying in number from 150 to 300, to show the relationship between ratings of different traits: Intelligence with personal qualities, with leadership, with physical qualities, and all the others with each other. I do not report these for I believe that the

TABLE XII.—MEDIAN, Q AND Q_3 OF TOTAL RATINGS GIVEN OFFICERS APPOINTED TO CAPTAINCIES FROM CIVIL LIFE WITHOUT PREVIOUS MILITARY EXPERIENCE, COMPARED WITH MEDIAN, Q AND Q_3 OF TOTAL RATINGS GIVEN CAPTAINS WHO HAVE BEEN FIRST AND SECOND LIEUTENANTS

	Captains, civil life			Other captains		
	Median	Q	Q_3	Median	Q	Q_2
Ordinance	67.5	60.64	74.85	67.7	60.9	75.9
Engineers	67.5	60.5	77.4	68.9	62.8	76.6
Quartermaster	71.01	63.51	78.7	71.9	64.3	80.9
Signal corps	68.5	59.8	72.4	69.5	62.6	77.8
Coast artillery	74.3	66.9	80.7	72.99	63.9	81.2
Infantry	75.48	65.48	81.74	75.12	66.78	82.92
Camp Taylor, October ratings	61.33	.....	.....	57.9		
Experimental group	69.8	.....	.....	61.4		

ratings were so inadequate as to make correlations computed from them of little interpretive value. Could I know the relative validity of the different judgments on a given person I am confident that a selection of ratings could be made and an average rating obtained which would be a valid measure of his traits. Unless that is done (as it is being done in a study we are now making in the Department of Educational Psychology of The Lincoln School of Teachers College) correlations between ratings should be ignored.

We have now reviewed the principal types of evidence that have been brought forward in connection with the practice of measuring our fellows' traits by means of single judgments obtained on point scales. Should we continue to place dependence on such measures of character? Most emphatically—NOT. We know of course from our general experience and from a study of the evidence I have set forth that *it is possible* to find raters whose discrimination is accurate and whose judgment of character will correlate very closely with objective

measures of it. Dr. Chassell found raters whose judgment correlated 0.7 with objective measures. But she found more whose judgments correlated 0.4 and 0.3 and 0.2 and 0.1 and 0.0. And the number of such is so large that we dare not use this method of measuring character, with the competency of raters as it exists today.

What, Then, Can We Do About It?

There are several things we can do about it. I name three and, I think, in order of increasing cruciality.

I. Average Several Judgments Made by Competent Judges

This is the first thing to do, by all means—increase the number of ratings on a person. *Obtain a mass judgment from good judges.* Discriminate carefully between judges and permit only perfectly competent people to rate. Assuming qualified raters, the reliability of a judgment increases directly with the square of the number of judgments. To double the reliability, take four times the number of judgments. The Probable Error of a single judgment is 0.6745σ; of two judgments it is 0.47σ; of three judgments 0.38σ; of four judgments 0.34σ.

Now the standard deviation of the best single rating of character we have yet been able to get in quantities on an 80 point scale is between 8 and 9 points. (As reported from this investigation the P.E. of a single rating is about 6 points.) Hence if it were possible to obtain four valid judgments on a person the P.E. of the average rating would be in the neighborhood of 3 points. The P.E. of the average of three ratings would be slightly more than 3 points. *This would insure that practically all averages of three or four judgments would locate a person within his proper fifth of the rating scale.*

But this presupposes that great care be taken to follow such a technique as was outlined on pages 485 and 486 of this report. In addition, I suggest that in ranking the persons from whom the scale-men are to be chosen, additional precautions should be taken to further objectify the process of ranking. Specifically, there should be made out for each person in the rank-list a "checking sheet" on which the subordinate elements entering into each group of qualities should be evaluated in 7 groups and a composite score entered for each person for that group of qualities. From these scores the rank order for each group of qualities will be of distinct value and the selection of scale-men very good indeed. I give next an illustration for physical qualities and intelligence to show the form of the suggested evaluation.

Checking Sheet

To be Filled for Each Officer Appearing on an "Original-list"

	1	2	3	4	5	6	7
	Very inferior	Inferior	Below average	Avurage	Above average	Superior	Very superior
Group Score 1–7							
I. Physical qualities:							
Physique							
Bearing							
Neatness							
Voice							
Energy							
Endurance							
Group Score 1–7							
II. Intelligence:							
Accuracy							
Ease in learning							
Ability							
(1) To grasp quickly the point of view of the commanding officer							
(2) To issue clear and intelligent orders							
(3) To estimate a new situation							
(4) To arrive at a sensible decision in a crisis							

II. Answering Questions about Traits of Character

Point scale rankings are made largely for general administrative purposes. They aid the parent and the school (or the teacher and administrator, if it is a teacher who is being rated) only very indirectly. The primary purpose of rating is diagnosis and improvement through conscious effort. An administrator wants to know definite facts about a teacher in employing or promoting him. "Is he loyal?" "Does he work well with others?" "Is he physically strong?" "Is he on the job?" etc. The parent wants the school to discover particular traits in his child and deal with those so definitely as to contribute to their improvement. To what extent: "Is he cheerful?" "Is he

honest?" "Is he shy?" "Is he conceited?" "Is he industrious?" "Does he stick to a job?" "Is he a chatter-box?" "Is he sensitive to beautiful things?" "Does he play fair?" "Does he work with the team?" and so on.

Now the school is responsible for discovering the answers to these questions about the pupil and the supervisory staff about the teachers. The experience of school people leaves no room for doubt that they *can* be answered—not in points with weighted credits, not in rank order or even in 7 or 5 groups. That refinement is not desired especially. After all the practical need is: In what traits is this person conspicuously deficient? Which ones should be consciously developed—which ones suppressed? For it is a fundamental postulate of the current educational order that the complex social and dynamic qualities can be made to respond to training. And the most effective method of laying one's fingers on the sore spots of a person's personality is to question keenly and thoroughly about it. We have already commented on the need for resorting only to thoroughly competent judges.

What one really wants is the child's personality profile, sketched through the answers to definite questions about particular phases of it. A very good picture, indeed, can be given of a child by merely the process of checking those questions, in a list of say fifty like the following, which the rater happens to be able to answer. Certainly the composite of three or four such judgments on a child would sketch his "personality," "temperament," "character," rather fully and helpfully. It would give a definite diagnosis and a practical lead for training. This tentative list of questions is now being answered about each pupil in The Lincoln School of Teachers College. The answers will be the principal basis of letters written to the parents of the children describing the child as we see him and suggesting methods of training in which school and home can cooperate.

Illustrative Questions to be Answered about a Pupil's Traits. Check + Traits above Average to a Marked Degree; Check − Traits Below Average to a Marked Degree

1. Is he cheerful?
2. Does he have a sense of humor?
3. Is he neat and tidy about desk and clothes?
4. Is he honest?
5. Is he dependable?
6. Is he unselfish?
7. Has he self-control?

8. Has he initiative?
9. Is he shy?
10. Is he a good loser?
11. Is he self-confident?
12. Is he conceited?
13. Is he careful with books, pencils, etc.?
14. Is he punctual?
15. Is he truthful?
16. Is he sensitive to criticism?
17. Does he take responsibility for his own acts?
18. Is he obstinate?
19. Does he excuse his own faults and mistakes?
20. Does he abuse privileges?
21. Does he demand more than his share of time and attention?
22. Is he sensitive to beautiful things?
23. Does he stick to a job until it is finished?
24. Does he use his leisure time advantageously?
25. Is he a bluffer?
26. Is he industrious?
27. Is he suppressed?
28. Does he consider the rights and feelings of others?
29. Does he cooperate?
30. Is he courteous?
31. Is he a snob? (Does he consider himself superior to others?)
32. Does he like to tease?
33. Is he rough in his play?
34. Does he lead on the playground?
35. Can he handle people well?
36. Does he take an active part in group activities?
37. Does he take his share in group activities?
38. Is he quarrelsome?
39. Is he interested in what others are doing?
40. Is he popular in his own group?
41. Does he obey school rules?
42. Does he respect authority?
43. Can he organize his ideas effectively?
44. Does he understand explanations and directions quickly?
45. Does he have ability to concentrate?
46. Does he work independently?
47. Does he apply his own experience and thought to the subject at hand?
48. Does he have good habits of work and study?
49. Does he ask intelligent questions?
50. Does he express his ideas well ?

A Second Method of Checking Pupils' Traits.—Another classification of traits and form of questioning is given herewith. This form is more definitional in character and specifies more carefully the situations on which the performances of the pupil are to be rated. It, too, implies simply a three-group rating. "Is he high or low—*i.e.*, conspicuous for the presence or absence of the trait?" If not ignore his record or check it "mediocre," "average" or what-not. Examples of this type of questioning for teachers and for students are given herewith. Either or both of these two cards have been in use in about 75 school systems since 1919. The more direct and simple form of questioning, referred to earlier is just being experimented within our school. Comparative data on the two types are not at hand now but should be before the end of the school year.

Certainly a definite recommendation can be made however:

Analyze specific qualities in your pupils or teachers. Aid your analysis by the direct answer to definite questions about particular performances. Don't attempt to refine the judgment by assigning points on a scale or classifying persons in 7 or 5 groups. Merely check those traits in which the person is conspicuous. Deal with his personality in terms of this diagnostic analysis. Get several judges to answer the questions on each person. Be positive that the judges are thoroughly competent to rate and let them rate only on those traits about which they are in no doubt.

This then is what I would do if I were compelled today to secure judgments on pupils or teachers. The occasion will be rare indeed when the point-scale rating will be needed in practical school work with pupils. In scientific work there will be frequent need for "point ratings." In such cases no rating should be used that is not the average of at least three independent ratings by competent judges and, in my judgment, made on a scale as objectified as the man-to-man comparison scale when constructed by the methods suggested.

III. Needed: A Scientific Analysis of Personality and Objective Tests of the Social and Dynamic Traits

Study the data of this investigation as you will, one conclusion presses insistently—the complex traits of character must be measured objectively, not judged. Certainly public schools will find it administratively difficult, if not impossible, to secure three independent competent ratings on either a pupil or a teacher. For diagnostic and training purposes helpful analyses *can* be made by question schemes like those outlined. But for the advancement of the science of education and the better fitting of people for their life work and play we need to analyze character in great detail and to measure it objectively. The measurement of the dynamic personal and social traits stands today where the measurement of intelligence did 15 years ago—on a purely subjective basis. Twenty years of laboratory measurement of specific functions preceded a decade of important synthetic work. Since Binet's original synthetic scale of 1908 we have moved rapidly forward in the objective measurement of general intellectual ability, through the betterment of individual tests and the initial construction of group tests for measuring intelligence.

But "personality," "character," "temperament," "disposition" —the fundamental complex attributes and their component dynamic traits—we have not even analyzed into their constituents, let alone

A Self-diagnosis and Improvement Chart

I. Skill in Teaching	Low	Aver.	High
To what extent:			
Does he know the subject matter of his own and related fields:			
1. In subjects like history, geography, etc., does he make effective use of material outside the text book			
2. Does he relate lessons to material in other fields and use illustrations outside his own subject (*e.g.*, mathematics and science)			
Does he select subject matter effectively for class reading and discussion			
Are his aims of teaching clearly defined			
Does he give evidence of having:			
1. Formulated clearly his aims of teaching, as shown by his written statement of aims and outcomes			
2. Planned his lessons specifically to carry these out			
3. Distinguished clearly between (*a*) "formal skill" (either in manual or academic subjects), (*b*) "information" and (*c*) "problem solving" as proper outcomes from his class work			
4. Given pupils clear ideas of the purposes of lessons			
Is he skillful in conducting the class discussion			
(*a*) Resourcefulness in organizing a discussion and in "thinking on his feet"			
1. Is he fertile and quick in taking advantage of pupils' questions			
2. Are his questions systematically planned, yet spontaneously given			
3. Does he express himself clearly			
(*b*) Skill in conducting "drill" exercises			
1. Does he make use of economical, "timed," drill-devices (such as Courtis' Practice Exercises, etc.)			
2. Does he properly subordinate drill to clear exposition; that is, keep a proper balance between drill and "development"			
(*c*) Ability to "develop" new phases of the work			
1. Are lessons well related to previous ones			
2. Is material "organized"			
3. Do lessons show the use of material in the solution of present or future problems:			
(*a*) In his subject			
(*b*) Outside his subject			
(*d*) Ability to secure class participation in the recitation			
1. Do all pupils in the class take part in the discussion			
2. Do the pupils question each other and conduct the class independently of his formal direction			
(*e*) Skill in making the assignment			
1. Was it an attempt to teach pupils how to study the lesson			
2. Was it more than mere formal announcement of the number of pages in the text, etc.			
3. Is its scope and purpose clearly recognized by pupils			
Has he insight into "how children learn"			
1. Does he keep the discussion within the pupils' comprehension			
2. Does he endeavor to discover pupils' difficulties by keeping records of errors and studying these			
3. Does he adapt discussion to individual differences in pupils			
Summary rating on skill in teaching			

II. Skill in the Mechanics of Managing a Class	Low	Aver.	High
To what extent—			
1. Does the class work proceed smoothly (without artificial interruptions and transitions from one kind of discussion to another)			
2. Do the pupils attend naturally and spontaneously to the work of the lesson			
3. Does order, or discipline inhere in the work (not maintained by compulsion nor suppression)			
4. Is routine, as passing material, moving to the blackboard, etc., economically and systematically organized			
5. Is material and equipment in the room effectively arranged			
6. Does he pay attention to the details of heat, light and ventilation			
Summary rating			

III. Team Work Qualities	Low	Aver.	High
To what extent—			
1. Does he cooperate with other teachers in school activities (committee work, Parent-Teacher Association, etc.)			
2. Does he contribute to faculty meetings			
3. Is he loyal to the administration and to other teachers			
4. Does he suggest plans for group improvement of the school			
5. Does he shoulder responsibility for his own acts			
6. Do pupils go to him voluntarily for advice and conference			
7. Does he go out of his way to advise and help students			
8. Does he acquaint himself with pupils' home conditions where it is wise			
9. Does he participate in community activities outside the school			
10. Are his records and reports in on time and in complete form			
Summary rating			

IV. Qualities of Growth and Keeping Up-to-date	Low	Aver.	High
To what extent—			
1. Does he read professional literature—books, journals, etc.			
2. Does he participate in and contribute to the discussion of educational meetings (teachers' association, etc.)			
3. Does he take extension courses, attend summer sessions, etc.			
4. Does he experiment with new methods in teaching which others have suggested			
5. Does he invent and experiment with new methods of teaching			
6. Does he heartily cooperate in investigational work in which various schools participate			
7. Does he participate on committees of associations in his own subject			
8. Does he contribute to educational literature			
Summary rating			

V. Personal and Social Qualities	Low	Aver.	High
To what extent—			
1. Does he attract people to him (*i.e.*, is he interested primarily in what others are doing)			
2. Does he meet people easily			
3. Does he recognize the importance of trimness in dress and general personal appearance			
4. Is he "fine-grained" (*i.e.*, is he sensitive to social proprieties)			
5. Does his impression of his own ability operate to handicap his effectiveness			
6. Is effectively aggressive in conversation and conference			
7. Is he tactful in dealing with pupils, colleagues and patrons			
8. Does he "eventuate," *i.e.*, does he carry through projects which he starts			
Summary rating			

Self-improvement Through Self-rating

To the Teacher.—Rate yourself on each quality on this form. It will be a first step in self-mprovement. It is important that you stand high in these qualities.

To the Principal or Superintendent.—Let the teacher rate himself on each question at least once each term. Self-analysis is the first step in self-improvement. To analyze human qualities well, one needs a definite and detailed guide. For effective teacher rating, both teacher and administrator should rate and confer on specific qualities which make for good teaching. A valuable file of the administrator's analyses of his teachers can be kept in the office.

make adequate tests for them. Probably the two objectives will be attained together—by a scientific analysis of the components of character and temperament we shall arrive at a tentative basis for making tests of the dynamic traits; by setting up tests and by correlating scores upon them with practical life criteria we shall refine our analyses of the complex products. It appears therefore that we should bend our efforts to making a scientific analysis of personality and to developing tests for the social and dynamic traits.

At least it helps to know that we are actually embarked on the adventure of trying to measure the more elusive characteristics of human personality. Crude beginnings have already been made; preliminary feelings around for the basic controls of conduct, likewise rough measures of their elements. Two illustrations can be considered briefly in closing this discussion—a discussion in which we have gradually been brought to the recognition that it is futile to depend longer on subjective estimates of character.[1] The first is an attempt to measure the constituents of will-temperament—to draw the "will-profile." The test appears in forms adapted to both individual[2] and group[3] testing. Ream has modified Downey's original test and has used it in prognosing the ability of salesmen. He finds a distinctly close relationship between total scores on the test and success as salesmen.

Downey is not interested in a composite measurement of "personality" or "temperament;" rather in an analytical "profile" of the person's component traits. An illustration will show what comes out of her testing. I have taken this from her "Manual of Directions."

Speed of Movement VI-1	Resistance to Opposition XII
Freedom from Load II-1, 2; VI-1, 2	Finality of Judgment XIII
Flexibility VIII	Motor Inhibition VII
Speed of Decision I	Interest in Detail IX
Motor Impulsion X	Coordination of Impulses V
Reaction to Contradiction XI	Volitional Perseveration VIII-2

Profile X. Profile X is that of a man who has held successfully a number of important executive positions. He is, in addition, an effective public speaker and possesses great dramatic talent.

[1] Downey, June E.: "Downey Individual Will-temperament Test." World Book Co., Konkers, N. Y.

[2] Downey, June E.: "The Will-temperament and Its Testing." World Book Co.

[3] Ream, M. J.: "A Group Will-temperament Test." (Test secured from author, Carnegie Institute of Technology, Bureau of Personnel Research.)

His profile suggests, in general, the type of the successful administrator, especially with reference to the high scores for speed of decision, finality of judgment, freedom from load, resistance to opposition, and motor impulsion in conjunction with high motor inhibition.

The high score for flexibility and the medium one on reaction to contradiction (tactful response) indicate social pliability and suggestibility which increase X's social assets, but are of dubious value in his business life.

The low score on interest in detail is not a serious defect, since X is in a position to turn over to subordinates the execution of many of his projects. It goes, however, with a tendency to generalize on insufficient grounds. The low score on volitional perseveration is probably a real weakness, although X's dramatic gift makes it possible for him to achieve through imitation what others work out through prolonged trial and error.

For those who do not know Miss Downey's conception of what the different tests measure, I quote a page from her manual.

"Speed of Movement: Speed of movement relative to the size of person and age; whether a person naturally moves quickly or slowly.

"Freedom from Load: Tendency to work at one's highest speed without external pressure; little tendency to relax speed; quickness in warming up to a task.

"Flexibility: Ease and success in readjustment; capacity to modify one's routine reactions. A very high score probably indicates some finesse in the handling of personal relations, or dramatic ability.

"Speed of Decision: Quickness in reaching a decision or conclusion. A slow reaction here may be due to caution or conservatism in weighing the elements involved in a situation or be caused by one's being sidetracked by irrelevant matters or by a rambling procedure.

"Motor Impulsion: Impetuosity and energy of reaction. The ease with which brakes or inhibitions are removed and also the tendency to an explosive reaction when the brakes are actually off.

"Reaction to Contradiction: This refers to the degree of confidence with which one maintains his opinion against contradiction. The reactions range from an aggressive attitude in which the burden of proof is thrown on the person who does the contradicting down to complete failure to assert one's own opinion.

"Resistance to Opposition: The vigor with which one reacts immediately to a blocking of one's purpose. It grades from a strenuous reaction, to complete passivity in the face of opposition.

"Finality of Judgment: Tendency to think a matter through and abide by one's decision. A moderate time may be given to revision in the interest of accuracy or as a provision against error in recording decisions. A low score characterizes an individual who keeps reopening a question and who shows vacillation in action since his judgment shifts with each shift in attention. My most extreme record comes from a man who in changing gears while driving an automobile wavers so long that the need for a shift is often over before he has decided what to do.

"Motor Inhibition: Capacity to keep in mind a set purpose and achieve it *slowly*. It involves power of motor control, imperturbability, and patience.

"Interest in Detail: Attention to details. This trait is not equivalent to accuracy, which usually carries an implication of power of keen analysis. One may possess great capacity for detail and yet lack penetration in the selection of details. Care for detail is more evident in execution of a plan than in cleverness in making a plan.

"Coordination of Impulses: Capacity to handle a complex situation successfully without forgetting any of the factors involved. This trait is probably allied to keeping one's head in a confusing situation, as in learning to drive an automobile, when the clutch, throttle, gear-shift, and brake must all receive attention.

"Volitional Perseveration: Absorption in a task; willingness to keep 'plugging away' at it because the examinee sets up a goal for himself."

In the Lincoln School of Teachers College we are carrying on a study of the dynamic traits and their measurement. In the course of it we are evaluating the existing tests, the Downey and Ream tests included. We will report the results of this evaluation later. It should be pointed out in passing that Miss Downey uses principally one means of measurement—motor coordination as exhibited through handwriting done under different conditions. To date Ream's brief study is the only one reported in which total scores on such tests are correlated with actual success in a life activity.

One other lead is being followed in measuring the dynamic traits—that of obtaining a direct record of actual performance in some practical life activity. Voelker's analysis[1] of the function of ideals and attitudes in social education and his tests for *trustworthiness* have supplied new examples. He has made a careful analysis of the rôle played by ideals and attitudes in the control of conduct. Taking the "ideal" or "trustworthiness" as the objective of a course of "scout" training, he measured results by a series of fourteen tests. "Trustworthiness" is measured by such tests as the following. Voelker's report shows that the scout training effected marked development in the growth of the ideal set up as the goal.

1. *The Puzzle Test.*—Can the subject be trusted not to steal an object which appeals to his interest and to his cupidity?

2. *Lost Article Test.*—Can he be trusted to make a sincere effort to return a lost article to its owner?

3. *Duck-on-the-Rock Test.*—Can he be trusted not to cheat in a game?

4. *Memory Test.*—Will he cheat in an examination?

5. *Overstatement Test.*—Will he refuse credit which is not due him?

[1]Voelker, Paul F.: The Function of Ideals and Attitudes in Social Education. Teachers College *Contributions to Education*, No. 112.

6. *Suggestibility Test.*—Will he stick to a point when he knows he is right?

7. *Let-Me-Help-You Test.*—Will he refuse help in solution of a puzzle when he has been instructed to solve it alone.

8. *"A" Test.*—Will he resist distracting interests?

9. *Profile Test.*—Can he be trusted not to peep when he is placed on his honor to keep his eyes closed? etc.

These examples of the measurement of the dynamic traits then, crude as they are, mark the beginning of a new stage in the analysis of the controls of conduct. Of two things the student of the scientific study of education can be sure: First, single point ratings of character are practically valueless; second, the fundamental social and dynamic traits play a large rôle in the control of the conduct of different individuals. They can be and are being measured. As Thorndike has well phrased the matter: "Whatever exists, exists in some amount." Let us go vigorously about the carrying through of the scientific analysis of "personality," "character," "temperament" and the objective measurement of their basic contributory traits.

TIME SAVING IN THE STANFORD-BINET TEST

EDWARD A. LINCOLN

Harvard University

The clinical examiner whose work requires that he examine large numbers of subjects either in the schools, an out-patient clinic, or an institution, finds himself continually pressed for time in the handling of his cases. It seems pretty well agreed that an hour is not too long for a complete examination with the Stanford revision, and if this is accompanied by some of the performance and school subject tests, as the best practice now requires, the time necessary for one individual is exceedingly long. To meet this difficulty several abbreviated scales have been devised,[1] but these have not yet found general favor among examiners. In the work of the Psycho-Educational Clinic in the Graduate School of Education at Harvard University there have been developed several points in the technique of administering the Stanford Revision which aid materially in reducing the time and effort necessary, and which this paper proposes to set forth.

If it can be said that there is one principle at the basis of the several points to be enumerated, this principle is the utilization of the "set" or mental attitude of the child, so that when he is started on one kind of test he is carried through as high as he can go before another field is attacked. For instance the repetition of digits occurs in the year III, IV, VII, X, XIV, and XVIII. When the directions for this test are first given to the subject in, say, year IV, it is very easy to go onto the longer numbers without intervening tests. This obviates the necessity of repeating the directions when a similar test is to be given in the upper years, and thus saves considerable time, especially if the range of subject is a wide one. It must be remembered, however, that, as Dr. Terman points out,[2] the younger children will, in many cases, not be able to give sustained attention to the same kind of a test for more than two or three minutes at a time. The practice in this respect must be guided by the age of the subject and the manner in which he

[1] Terman gives a short scale in connection with the Stanford Revision. An abbreviation almost identical with this was used in the army. See also *Psychological Clinic*, Vol. XI, 1917–18, p. 210, "A Brief Binet-Simon Scale," by E. A. Doll. Also, *School and Society*, Vol. X, No. 259, Dec. 13, 1919, "An Abbreviated Mental Age Scale for Adults," by Lincoln and Cowdery.

[2] "The Measurement of Intelligence," p. 196.

is responding. If he succeeds easily with the first set of numbers in any group, it is undoubtedly safe to go on to the next higher year in which the test occurs. When, however, there are signs of drooping interest and lagging attention it is better to shift quickly to some other field.

The method of giving tests in groups has another advantage. After one or two such groups have been given, it is usually possible to place the subject very accurately on the scale, and thus save time by determining at the outset the upper and lower limits beyond which he is not likely to go. It is our experience that beginners in testing waste considerable time in giving tests which are either much above or much below the abilities of the subjects with whom they are working.

Among the groups or series of tests which we have found most useful are the following:

Comprehension	Similarities	Digits Forward	Vocabulary and Definitions
IV, 5	VIII, 4	III, alt.	V, 4
VI, 4	XII, 8	IV, 6	VIII, 5
VIII, 3		VII, 3	VIII, 6
X, 5		X, alt.	X, 1
		XIV, alt.	XII, 1
		XVIII, 3	XIV, 1
			XVI, 1
			XVIII, 1

Drawing	Digits Backward	Repeating Sentences	Differences
VI, 4	VII, alt.	III, 6	VII, 5
VII, 6	IX, 4	IV, alt.	XIV, 3
VIII, 1	XII, 3	VI, 6	
X, 3	XVI, 5	X, alt.	
XII, 4	XVIII, 5	XVI, alt.	

Some further points may be noted in connection with the use of these groups. It has been found advisable to give the "digits backward" tests before the "digits forward," and that the two series should be separated by other tests. This is because the latter is the more natural reaction, and thus a good bit easier. If it is given first it is sometimes impossible to break up the "set" of the subject so that he can give the digits backward. A similar consideration holds in the case of the similarities and differences. The latter seem to be much easier, so the similarities should be given first, and several tests should intervene before the differences are given.

The advantage in the drawing series is that once the child is given the pencil he will do all that is to be done with it at once, thus saving

the time which is ordinarily consumed in giving him a pencil and taking it away several times.

Much time may be gained or lost at the beginning of the examination when the examiner is getting into the good graces of the child. Most beginners make too lengthy and ponderous a business of getting "rapport." In almost every instance if the child is greeted with a smile and asked if he would like to do some puzzles there will be little difficulty. Much depends on the test with which the examination is opened. We make it an almost invariable rule to begin with the pictures. Nearly every child is interested in them, and will make some response to them. Furthermore, the picture test is a great help for the preliminary placing of the subject, for it receives credit in year III, VII, or XII according to the kind of reaction.

Let us see how this scheme would work with a typical case. Suppose a nine year old subject is given the examination. He is shown the pictures, and describes all except one, in which there is a little interpretation. Thus he scores plus in VII, 2, but minus in XII, 7. We then pass to the group of comprehension questions, beginning in the middle of the series with VIII, 3. He passes this, and also, X, 5. The next series will be the memory for digits backward, in which he gets four digits in IX, 4, but misses the five at XII, 6. The vocabulary test should come next, and in this he gets 32 words, thus passing VIII, 6 and X, 1, but failing XII, 1. It now has become reasonably clear that the subject's mental age is somewhere near the nine or ten year level. The remaining tests at these years should now be given, further exploration being unnecessary.

It is, of course, highly desirable to get as complete a record of the child as possible. However, it becomes necesary at times to sacrifice somewhat in thoroughness for the sake of saving time. In the tests where a number of responses are required it is unnecessary to go on giving the various items after the subject has failed in so many items that he cannot possibly receive credit for the test. In the case of the Absurdities (X, 2) for instance, after two have been missed the test cannot be passed, so it may be marked immediately with a minus sign, and may be left for something else. Other tests in which this procedure may be used are found in every year.

It is also unnecessary to go on giving further items in a test after the subject has done enough correctly to give him credit. In the Definitions test at VIII a child has to give only two out of the four required definitions in order to receive credit. If he responds cor-

rectly to the first two there is nothing gained by giving him the third and fourth, and if he gets two out of the first three there is no need to give him the last one. A case of this sort where considerable time is likely to be saved occurs at XIV, 3, the differences between the president and the king. Many children get two of these differences immediately, but cannot find a third, or can discover it only after long study.

The use of the tests in this manner requires that the examiner be thoroughly conversant with the location of the various tests, and that he be able to score the responses without reference to the directions, except in the occasional case of the uncommon reaction. It is also absolutely necessary, if the full benefits of this method are to be gained, that the materials for the test be arranged in a convenient way and that the examiner be completely familiar with this arrangement.

A YEAR OF THE EDUCATIONAL RESEARCH COMMITTEE

SAMUEL P. CAPEN

American Council of Education

The Educational Research Committee of the Commonwealth Fund held its first meeting approximately a year ago. Its members believe that the educational public will be interested in a brief report of the transactions of the Committee and of the educational research now going forward for which it stands sponsor.

In the summer of 1920 the Commonwealth Fund, at the suggestion of Professor Max Farrand of Yale University, then the Fund's General Director, appropriated $100,000 for the purpose of encouraging educational research. It was understood that if satisfactory results were obtained from the expenditure of this amount during a single year, similar appropriations would be made annually for a period of five years. The policies to govern the expenditure of the appropriation were left to later determination.

The General Director organized a Conference of persons experienced in conducting or directing educational research, which met for three days in October, 1920, and recommended a plan of procedure to the Directors of the Commonwealth Fund. The plan proposed a departure from the current practice of philanthropic foundations in the conduct of educational research. Instead of setting up a more or less permanent agency with an expert personnel, it was recommended that the Commonwealth Fund subsidize individual investigators of proved capacity or of great promise to undertake limited researches. The Conference further indicated certain large fields in each of which numerous painstaking scientific studies are needed. These are: School revenues; the evaluation of school subjects and the determination of standards of accomplishment in them; reorganization of the administrative units of the public educational system; the establishment of standards and methods of supervision. The Conference also recommended that the Commonwealth Fund appoint a committee to consider and recommend projects for research and to assume executive responsibility for supervising the carrying on of such researches as might be subsidized by the Fund.

The Directors of the Commonwealth Fund accepted the Confer-

ence's recommendations and appointed as the Educational Research Committee, Leonard P. Ayres, Samuel P. Capen, Lotus D. Coffman, Ellwood P. Cubberley, Charles H. Judd, Paul Monroe and Frank E. Spaulding. Professor Max Farrand, the General Director of the Fund, was designated to act as Chairman. Since the organization of the Committee Professor Spaulding has been obliged to resign and President James R. Angell has been appointed in his stead. During Professor Monroe's absence in the Orient, his place has been taken by Professor E. L. Thorndike. Professor Farrand has resigned as General Director of the Commonwealth Fund but remains as Chairman of the Educational Research Committee.

The Committee's general policy has followed closely the lines of the recommendations made by the Conference above referred to. During the year in which it has been in existence, a considerable number of requests for subventions have been presented to it. These have been exceedingly varied. Some of them have come from persons of no reputation as investigators and have been very vaguely defined. Some have been presented by distinguished scientists but called for the support of investigations which could hardly be classified as educational research. Certain requests have been made for the subsidization of special departments or of individuals in colleges or universities, without specification of the research projects to be supported by the subsidy. Other requests submitted by persons of known competence have sought subventions for projects carefully defined and budgeted. After a preliminary review of these heterogeneous askings, the Committee came to several conclusions which have since met with the approval of the Commonwealth Fund. In the first place, it decided to recommend no subventions to departments or individual workers in institutions for the carrying on of the regular research activities of such departments or individuals. Secondly, it determined to recommend the support of only those projects which were carefully defined both as to objectives and as to methods and which were accompanied by an itemized estimate of the cost of the undertaking. Thirdly, it decided for the present to recommend no subsidy for a longer period than one year. Within that time the investigation must either be terminated or a substantial report of progress submitted. Fourthly, the Committee recommended that wherever possible the Common wealth Fund should have its financial dealings with the institution or organized agency to which the investigator is attached, rather than with the individual.

Since this last mentioned policy of the Committee has aroused considerable interest in various quarters, the form of contract which the Committee has devised is here quoted:

The institution will accept grants for educational researches from the Commonwealth Fund and will be responsible for their disbursement under the following agreements:

1. Salaries of officers who are relieved of regular duties to engage in researches are to be charged against the research grants at the rate of the salaries paid by the institution to such officers for regular teaching and administration, except in cases where explicit exceptions are arranged in advance.

2. The institution will disburse the grants under the following arrangements: On acceptance of the grant by the institution, the Commonwealth Fund shall deposit with the business officer of the institution a sum suitable to launch the investigation and determined on the basis of the size of the grant; in the case of large grants this sum will amount in general to 20 or 30 per cent of the grant. When the initial sum is approaching exhaustion the business officer of the institution shall request a second deposit and shall render, as soon as possible, a full account of expenditures of the first deposit. In this manner there shall be successive deposits and successive accountings of the grant until the total amount has been used.

In disbursing the funds the institution will assume administrative responsibility for all payments of salaries. It will approve all appointments of assistants. It will make payments on the order of the investigator for supplies and equipment, and traveling expenses, and will render accounts on the latter items, showing the approval of the investigator.

At the termination of the grant it is understood that any unexpended balance shall revert to the Commonwealth Fund, that final disposition of such supplies and equipment as are at hand is subject to the order of the Commonwealth Fund. If at the time of settlement property of any kind is left at the institution, it is understood that it becomes permanently a gift to the institution.

If the grant is made with specifications as to the amounts which are to be used for salaries, traveling expenses, and supplies, the institution will limit all expenditures to the classes of items specified and will allow transfers from one class to another only on explicit permission of the administrative authorities of the institution, but it is understood that readjustments within a single class of expenditures may depart from the original terms of the budget.

3. The person responsible for the investigation will be required to file a report on the investigation both with the administrative officers of the institution and with the Directors of the Commonwealth Fund at stated intervals.

It will be noted that the Commonwealth Fund does not propose to pay a bonus to persons who undertake educational research at its expense. The salaries paid investigators are to be the same as the salaries they would receive from the agencies which employ them. The Commonwealth Fund merely makes it possible for an investigator to carry on particular studies in which he is especially interested, and

if necessary to be temporarily relieved of his regular institutional duties without pecuniary loss.

The Educational Research Committee has held three regular meetings. Two of these were devoted to the assignment among the most promising projects of the appropriation made for the academic year 1920–21. At the third meeting held in October, 1921, a portion of the appropriation for the academic year 1921–22 was assigned. A brief account of the projects which have been supported may be of interest. It will be noted that these all fall within the first three of the major fields of study indicated in the initial report of the Conference.

Educational Finance

The Commonwealth Fund has joined with three other educational foundations in appropriating to the American Council on Education a sum sufficient to carry forward a comprehensive investigation of educational finance in the United States. The program for public education laid down in legislative enactments and state constitutions will be examined to determine to what extent communities are already meeting the public desires. Effort will be made to investigate the cost of the program designated by the public. The possibility of effecting economies will be studied. The relation of educational expenditures to expenditures for other governmental purposes will be worked out. Intensive studies will be made in individual states that may be regarded as typical and the most important facts covering the country as a whole will be assembled and collated. The American Council on Education has appointed a special commission to take charge of this investigation.

An appropriation has also been granted to Columbia University for the preparation under the direction of Professor George D. Strayer of an initial report on city school budgets.

Measures and Standards of Achievement in School Subjects

Appropriations have been made to Columbia University for the conduct of two investigations under the direction of Professor E. L. Thorndike. The first investigation deals with the possible reorganization of the teaching material in Algebra and the methods of presenting that subject. What is known about the psychology of Algebra is to be collected, gaps in that knowledge are to be noted and filled by appro-

priate investigations so far as possible, especially such as are important in possible changes in curricula and methods.

The second investigation relates to vocational guidance. It is designed to prepare standard tests of ability to continue school work, of ability to learn to do clerical work, and of ability in the mechanical trades and factory work. These tests are to be for use with boys and girls of approximately fifteen or sixteen years of age. It is expected that they will be so formulated as not to require the services of a psychologist to give them.

Two appropriations have been made to the University of Chicago, one for the use of Professor Judd and assistants in conducting a laboratory study of reading, and the other to Professor Morrison for devising a series of tests designed to measure the progress of pupils in French under ordinary high school instruction. In the investigation of reading, laboratory methods are used which teachers cannot employ. The movements of the eyes of adults and children are photographed under different conditions while they are reading various kinds of passages. It is expected that in this way the processes involved in good and bad reading and in mature and immature reading may be determined. Once the characteristics of various kinds of readings are ascertained it is possible to turn over to teachers many useful suggestions about the handling of pupils.

The French investigation is designed to throw light upon the effectiveness of grammatical as compared with non-grammatical methods in learning to read the foreign language; the pupil's command of grammatical usage in functional form compared with his knowledge of grammatical principles abstractly stated; and the relation between the ability to get the meaning of a series of French words stated apart from any context and the ability to react to the meanings of the same words when they are included in a piece of discourse.

An appropriation was made to be spent by the Chairman of the Educational Research Committee on a preliminary conference on the social studies. The conference outlined the problems in the reorganization of teaching material in the social studies and on the basis of its report the Committee has recommended further appropriations for a historical review of the social studies and an evaluation of current experiments in new methods of presenting these subjects.

An appropriation has been made to the Board of Education of Winnetka, Illinois, for the conduct of a study under the direction of Superintendent Carleton W. Washburne of periodical and reference

literature to determine the commonly known and referred to historical and geographical material, with a view to the possible reorganization of the school material for teaching these subjects.

A grant has been made to Leland Stanford Junior University for a study, under the direction of Professor L. M. Terman, of gifted children in California. At present such children remain unidentified and submerged in the school's masses. The usual curriculum methods leave their intellectual and volitional resources largely undeveloped, sometimes possibly perverted. It may be more important to discover and to give appropriate educational opportunity to a single gifted child than to prevent the birth of a thousand feeble-minded. The investigation proposes to secure certain basic facts with reference to approximately one thousand school children of exceptionally superior intellectual ability, and to follow up the records and achievements of those pupils over a period of years.

A subsidy has been granted to the New York Association of Consulting Psychologists for a study partly similar in its objects to that of Professor Terman's. It is proposed to give intensive psychological examinations to students in a group of public schools in New York in order to determine the ability of children as they enter school, classify them as to ability and follow them up by re-examinations and through the services of a home worker, and thus to lay the basis of possible modifications of courses of study for the benefit of intellectually superior children, and that the less able children may be given better opportunities for development.

Reorganization of the Administrative Units of the Public Educational System

The Fund has made a grant to the University of Minnesota to be under the direction of Professor L. V. Koos in studying and critically evaluating the present status of the junior college movement. There are now upwards of 300 of these institutions and they are multiplying rapidly. It is the purpose of the study to show their relations to secondary education, to the prevailing four year college of liberal arts, and to professional education. Such a study it is believed should have large influence in determining the trend of future efforts toward educational reorganization at the level of the lower years of the college course.

The Educational Research Committee believes that there should be

many more appeals for subventions than have thus far come to it and that requests should be made by a much wider range of institutions. Indeed the conditions of the grant and the policy of the Committee are so flexible that any first-class project which can be clearly defined and budgeted is likely to receive favorable consideration. The Committee meets three times a year, in the fall, in the early spring, and in the early summer. The next meeting will be held March 4, 1922. Projects to receive consideration must be in the hands of the undersigned at least two weeks before the meeting of the Committee.

EDUCATIONAL PSYCHOLOGY AT THE PRINCETON MEETING OF THE AMERICAN PSYCHOLOGICAL ASSOCIATION

ARTHUR I. GATES

Teachers College, Columbia University

Despite the fact that the visiting psychologists were divided between the session of Sections I and Q of the A. A. A. S. at Toronto and the Thirtieth Annual Meeting of the American Psychological Association at Princeton, at least 150 members appeared at the latter convention to participate in a thoroughly profitable session. In contrast with the meeting of a year ago, which was characterized by evidences of unrest, the Princeton meeting, as the result, probably, of the contentment following readjustment during a year of productive work, was marked by a cooperative good will and an enthusiasm for the solution of many of the traditional problems of the fundamental sort.

The outstanding feature of the meeting was the special session on "Psychology in its Social Relations," at which representatives of the medical, psychiatrical and psychological sciences freely discussed their mutual problems and misunderstandings. At this meeting, papers were read by R. C. Cabot (Harvard), S. Paton (Princeton), S. I. Franz (Government Hospital for Insane, Washington, D. C.), and C. M. Campbell (Boston Psychopathic Hospital), followed by informal discussion lead by W. McDougall (Harvard), F. L. Wells (Boston Psychopathic Hospital), R. S. Woodworth(Columbia), and C. E. Seashore (Iowa). The session on "Abnormal Psychology" was a continuation of the exchange of opinions among workers in these related fields.

Half-day sessions were arranged under the following titles: (1) General and Experimental; (2) Clinical; (3) General; (4) Mental Measurement; (5) Experimental; (6) Psychology in its Social Relations; (7) Industrial and Educational; and (8) Abnormal Psychology. There was considerable overlapping in the content of the several programs. Of the fifty papers presented, twenty-seven bore rather directly on specific problems of interest to Educational Psychology; of these, thirteen were concerned with tests or measurements, ten with learning, and four with the effects of fatigue, drugs, etc. on mental or motor efficiency.

There was abundant evidence in the content of formal papers and in the informal discussions that psychologists are keenly aware of momentous deficiencies in hypotheses that have been regarded, for purposes of application, as established principles. The warnings voiced with vigor at the Chicago meeting a year ago were obviously expressions of deeply rooted conviction. Generally speaking, the warmest approval, at Princeton, was given to those papers which presented investigations seeking for evidence on certain persistent problems of fundamental import.

Papers Dealing with the Psychology of Learning

Among the studies of a fundamental problem whose relation to Educational Psychology is intimate, was one described in a paper by Warner Brown of the University of California. A large group of subjects practiced a variety of functions over a period of thirteen weeks. The results of this study disclosed a striking inadequacy of our technique for measuring improvement and the vagueness of our knowledge of the general mechanics of learning. While the correlation between initial and final status in a function is positive, many exceptional cases of a striking character were found and marked irregularities in the course of improvement suggested the need of more refined analysis. Improvement in one function does not generally indicate a similar improvement in other functions, nor at other stages in the same function. In a similar study, G. S. Gates (Barnard) found that improvement was extremely variable; that improvement over the first half of the practice period is not substantially correlated with improvement during the second half of the practice period, with initial or final status, in the same function, nor with improvement in others. Final ability in one function gave correlations of about 0.4, with final ability in others, these correlations being considerably higher than correlations between initial scores in the several functions. W. S. Hunter, University of Kansas, found correlations from 0.17 to 0.45 depending on the measures adopted, between ability in a pencil maze and scores on the Otis Intelligence test, whereas, in the case of rates, performances in maze tests were so variable that such a concept as "general learning ability" could not be justified. Such studies disclose the inadequacy of our knowledge concerning the capacity to learn, the uncertainty of the principles upon which educational guidance has been conducted,

and the necessity of more thorough research in the whole field of "practice."

That emphasis upon speed rather than on accuracy results in the most expeditious and effective learning was the thesis of a paper by Grace E. Bird, Rhode Island State College. "Rapid drill from the beginning 'focalizes' and initiates habit without superfluous behavior." This fact was said to obtain in the case of certain industrial processes, in adding and in reading. This is probably a matter in which generalization would be risky; and the need of research in each of the various school functions is suggested.

The effect of motivation in the form of a wage bonus on the improvement of abilities among hand compositors as reported by H. D. Kitson, Indiana University, was an increase in output of sixty-seven per cent. The results of this study parallel the outcome of measurements of school subjects in which the comfortable mediocrity of efficiency in reading, writing, etc. usually found in the later grammar grades may be greatly surpassed as the result of the provision of an incentive to improvement.

A. S. Edwards, University of Georgia, found that instruction of school children in methods of study resulted in improvement in their work. Such instruction must be specific and apply directly to the tasks then being undertaken in the school room. A course of study in how to study is being constructed for use in the grades. This worker quite justly asserts that more than mere external motivation is essential to effective learning. Definite knowledge of what to learn, and how to learn it, is needed.

A. I. Gates, Teachers College, reported a study, the purpose of which was to analyze reading and spelling into their constituent elements, to devise a technique for the diagnosis of backwardness in these functions, to ascertain the causes of such backwardness and to try out certain types of remedial treatment.

Papers on Mental Tests and Measurements

The continuation of interest in the field of mental testing was evidenced by the fact that nearly one third of the papers dealt with measurement. The number of new tests presented was smaller than usual but the interest displayed in the critical evaluation of instruments now available was widespread.

L. M. Terman, Stanford University, described an extensive project now under way in California, for the discovery and study of approxi-

mately 1000 children of very superior mental ability. The study will embrace measurements of the important mental, physical, social, and temperamental traits, as well as a thorough survey of educational attainments, heredity, home surroundings, health, etc., and all told, promises to be the most extensive and thorough study of genius ever undertaken.

Bird T. Baldwin presented data concerning the relation between mental and physical growth based on consecutive measurements of individuals; a product of the admirable research, which is being conducted at Iowa by Baldwin and L. Stecher. These workers have found it possible to predict the stature at sixteen years of age from the measurements secured at ten with a PE of approximately 2.5 centimeters. The IQ can be predicted over a similar period with a PE of estimate of 6.3 points. In general, the curves of growth for physical and mental traits are very similar. The importance of physiological development in its bearing on mental, social and educational achievement, was stressed.

A new and expeditious method of computing multiple correlations was described by H. A. Toops, of the Institute of Educational Research of Teachers College, together with the general technique to be employed in the construction of scales for general occupational groups as contrasted with scales for specific vocations.

The several papers just mentioned disclose a situation in the progress of research which marks a new era; the organization of institutions whose personnel and equipment are entirely devoted to researc While the progress of the last two decades achieved by workers whos main task—that of instruction—has been great, the present decade with its organizations equipped wholly for research promises an unprecedented accomplishment.

A comparison of superior duplicate twins (IQs 183 and 181) by Arnold Gesell, Yale University, was a most interesting and convincing illustration of the infinite detail with which heredity may operate. The striking similarity of physical characteristics, even to the identity of slight peculiarities of teeth, of skin pattern or the appearance of a small mole, was paralleled by the correspondence of the results from a battery of educational and mental tests. This paper gave an admirable example of the thoroughness with which human traits may now be measured when the instruments of several sciences are employed.

A number of studies of the predictive value of tests or of groups or individual differences disclosed by them will be briefly summarized.

A. M. Gordan, University of Arkansas, correlated four well known tests of general mental ability with several criteria, finding some of them so specialized in their predictive value that particular exercises often gave higher correlations, *e.g.*, with arithmetic, than the whole scale. The need of tests for native ability in each of many school and other functions was suggested. David Mitchell, New York City, gave results obtained from the measurement of the general mental ability of 1000 children of pre-school age. Ada H. Arlitt, Bryn Mawr College, found a slight superiority in IQ of white over negro children which increased with age. Joseph Peterson, Peabody College, employing an ingenious multiple choice test, found a marked superiority of white over negro children, particularly in the scores representing higher mental operations.

C. S. Yoakum, Carnegie Institute of Technology, found certain items of the Downey Will-temperament test as modified by M. J. Ream, to be indicative of success in salesmanship, whereas tests of general mental ability, aside from assigning a minimum essential, had little predictive value. That certain specific reactions to ethical discrimination tests selected from the Stanford-Binet, are of value in predicting susceptibility to delinquency, was stated by Augusta F. Bronner of the Judge Baker Foundation.

Laura M. Chassell, Ohio State University, found that grades received in the preliminary examination for the degree of Ph. D. and ratings of the Doctorate thesis gave mean correlations of approximately 0.6 with various criteria of success in later work. Judgments based on letters of recommendation gave correlations with success ranging from 0.01 to 0.70, depending on the author of the letters. Correlations between moral traits and general mental ability, both determined by judgments of acquaintances, average approximately 0.5 according to an extensive investigation among college students by Clara F. Chassell, Teachers College.

Studies of the Effects of Fatigue, Drugs, Etc.

H. L. Hollingworth, Columbia University, in the course of an extended investigation of the effects of alcohol, discovered significant facts concerning the relation of proficiency to the susceptibility to the damaging effects of the drug which was suggestive of a promising line of research in pharmaco-psychology. It was found that those subjects who were most proficient in the tasks at the start, and those who improved the most during the practice showed the least susceptibility

to the drug. Since the various functions used constitute the equivalent of a measure of general mental ability, the implication is that the more intelligent adults have a superior general bodily equipment—"quality of the organism"—which is not only more adaptable to the environment in a general way but to such specific influences as alcohol.

Three papers on aspects of fatigue were read. F. C. Dockeray, Ohio Wesleyan University, constructed an apparatus after the pattern of the Dunlap low oxygen tests used in the Air Service, which betrays the periods of diminished attention occurring in states of fatigue. Buford Johnson, Johns Hopkins University, employed tests of the sugar content of the blood and urine as checks in an investigation of mental and motor work. The results were not conclusive. Florence R. Robinson, University of Chicago, persuaded a group of students to go without sleep for thirty hours, some of them for forty-eight hours, for the privilege of taking a series of mental and motor tests at intervals. The loss of efficiency was no greater than the amount of gain due to practice although "feelings of fatigue" were reported.

The address of the President, Margaret Floy Washburn, following the Annual Dinner, was an able defence of "Introspection as an Objective Method." This address will be printed in full in the *Psychological Review.*

The most important transaction of the Annual Business Meeting was the adoption of policies with regard to the technique of issuing licenses as "consulting psychologists" and the determination of qualifications for such licenses. The recommendations of the committee appointed in 1920 were, in all essentials, adopted.

The following officers were elected for the year 1922. President: Knight Dunlap, Johns Hopkins University. Members of the Council: Warner Brown, University of California, and F. L. Wells, Massachussetts General Hospital. Representatives of the National Research Council: J. McKeen Cattell, and E. G. Boring, Clark University. Twenty-three were elected to membership in the Association.

The meeting in 1922 will be held in Boston. F. L. Allport, Harvard University, was elected local representative of the Executive Committee.

PROGRAMS OF COMING MEETINGS

TENTATIVE PROGRAMS ARRANGED FOR OPEN MEETINGS OF NATIONAL ASSOCIATION OF DIRECTORS OF EDUCATIONAL RESEARCH, TUESDAY, WEDNESDAY, AND THURSDAY AFTERNOONS, FEBRUARY 28, MARCH 1, MARCH 2, IN THE GOLD ROOM OF THE CONGRESS HOTEL, CHICAGO

I. Tuesday. Mr. Harold Rugg, Presiding

PROGRAM OF RESEARCH IN MENTAL AND EDUCATIONAL MEASUREMENT

1. Results Obtained by Classifying 2000 Kindergarten Children by Means of the Binet Test. Charles D. Dawson, Public Schools, Grand Rapids, Michigan.

2. Research vs. Propaganda in Visual Education. Frank N. Freeman, University of Chicago.

3. Evaluation of Group Intelligence Tests. Raymond Franzen, Public Schools, Des Moines, Iowa.

4. A Study of Reading and Spelling with Special Reference to Disability. Arthur I. Gates, Teachers College, Columbia University.

5. The Anticipation of Meaning as a Phase of Reading Ability. C. T. Gray, University of Texas.

6. The Development of Certain Types of Reading Habits. Guy T. Buswell, University of Chicago.

7. Intelligence and Progress Through the Grades. Arthur W. Kallom, Boston Public Schools.

II. Wednesday. Dr. Lotus D. Coffman, Presiding

PROGRAM OF RESEARCH ON THE CURRICULUM AND SCHOOL PROGRESS

1. Curriculum Construction in an Experimental School. Otis W. Caldwell, Lincoln School of Teachers College.

2. Comparison of Reading, Writing and Pre-school Spoken Vocabularies. Ernest Horn, University of Iowa.

3. The Collection of Unrecorded Subject Material. W. W. Charters, Carnegie Institute of Technology.

4. Relation of Measurement to Pupil Progress and Curriculum Research in Reading. Laura Zirbes, The Lincoln School of Teachers College.

5. Temperament and Attitude as Factors in School Progress. Clara Schmitt, Bureau of Child Study, Chicago Public Schools.

III. Thursday. Charles E. Chadsey, Presiding

PROGRAM OF RESEARCH IN SCHOOL ADMINISTRATION

1. A New Supervisory and Administrative Organization for Public Schools. P. C. Packer, University of Iowa.
2. Educational Measurement as a Key to Individual Instruction and Promotions. Carleton W. Washburne, Superintendent of Schools, Winnetka, Illinois.
3. Qualities Related to Success in Elementary School Teaching. Frederic B. Knight, University of Iowa.
4. The Sociological Character of the Secondary School Population. George S. Counts, Yale University.
5. Methods of Investigation in the Field of Educational Finance. George D. Strayer, Teachers College, Columbia University; Director of Educational Finance Inquiry.

FOR MEMBERS ONLY.

TWO MEETINGS OF THE ASSOCIATION WEDNESDAY MORNING AND THURSDAY MORNING, MARCH 1 AND 2. ROOM TO BE ANNOUNCED, PROBABLY IN CONGRESS HOTEL

I. Wednesday morning meeting. General Topic: A Clearing House of Educational Research Now under Way throughout the Country.

Informal five minute talks will be made presenting succinctly examples of research in all fields of education; mental and educational measurement, curriculum studies, learning investigations, classification of pupils, school finance, school buildings, programs, promotions, etc. The president desires to receive from each member of the Association a brief outline of the research which he will report at this meeting. We should have not less than 20 such reports. Persons have been designated in the different research and training institutions to make reports of research under way in these places. The meeting is organized for the purpose of acquainting us with what our colleagues are doing, to clear our minds as to the direction in which we are moving and to set forth the strength and weaknesses of our present research practice.

II. Thursday morning meeting. The second closed meeting will deal with the preparation and publication of products of educational research. Eight or ten papers and reports will be prepared to discuss crucial issues of educational writing. The purpose of the conference is two-fold: (1) the improvement of educational writing; and (2) the encouragement and stimulation of research workers to publish in effective channels, and in appropriate form, the results of their research.

For Members and Invited Guests

III. The Annual Dinner, 6.30 Thursday evening. Place to be announced later.

1. Presentation of two honorary members elected at the last meeting: Dean James E. Russell, Teachers College, Columbia University, and Dr. G. Stanley Hall, President Emeritus of Clark University.

2. Annual Address of Retiring President: The Methods of Science in Educational Research. Harold Rugg, The Lincoln School of Teachers College.

PROGRAM OF SOCIETY OF COLLEGE TEACHERS OF EDUCATION

CHICAGO MEETING—1922

First Session—Monday, February 27, 2:30 P.M.

Progress and Present Status in the Scientific Study of Education

1. Mental Tests. S. S. Colvin, Brown University.
   Discussion led by F. N. Freeman, Chicago University.
2. Statistical Method. H. O. Rugg, Columbia University.
   Discussion led by ——————————
3. Subject Tests. B. B. Buckingham, Ohio State University.
   Discussion led by ——————————
4. Educational Determinism. W. C. Bagley, Columbia University.
   Discussion led by ——————————

Second Session—Tuesday, February 28, 2:30 P.M.

College Instruction in Education

1. The Place of the Project Method in College Courses in Education. W. H. Kilpatrick, Columbia University.
   Discussion led by ——————————
2. The Case for the Case Method. L. O. Cummings, Harvard University.
   Discussion led by H. Updegraff, University of Pennsylvania.
3. The Needs of the Educational Practitioner. Raymond W. Sies, University of Kentucky.

Discussion led by ——————————————

4. Business Meeting.

THIRD SESSION—Wednesday, March 1, 2:30 P.M.

THE ORGANIZATION OF COLLEGE DEPARTMENTS OF EDUCATION

1. The Distribution of Functions of College Departments of Education and of Normal and Training Schools. J. W. Withers, New York University.
   Discussion led by Livingston C. Lord, Charleston, Illinois.
2. The Relations of Departments of Education to Other Departments of the College or University. M. E. Haggerty, University of Minnesota.
   Discussion led by R. M. Ogden, Cornell University.
3. Standards for Professional Approval. W. S. Gray, University of Chicago.
   Discussion led by ——————————————
4. The Relations of College Departments of Education to State and City School Systems. G. D. Strayer, Columbia University.
   Discussion led by ——————————————

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY CECILE COLLOTON

Department of Educational Psychology, The Lincoln School of Teachers College

Educational Tests

The Minnesota English Composition Scales; Their Derivation and Validity M. J. Van Wagenen. Educational Administration and Supervision, 1921, December 481–499. Description of composition scales for exposition, description, and narration, evaluated independently for thought content, sentence and paragraph structure, and mechanical perfection. Data given to prove marked degree of stability between scales.

The Use of Educational Measurements in the Training Department of the State Normal School, Ellensburg, Washington. Mary A. Grupe and Elsa M. Smith. Educational Administration and Supervision, 1921, December 517–526. The efficiency of student teachers as shown by results of standard tests. Follow-up work in careful readjustment of instruction and grouping on the basis of tests.

The Quality of Freshman Composition. G. C. Brandenburg. School and Society, 1921, December 17, 579–584. Scoring freshman composition at Purdue University by five judges and by the Hillegas Scale. Comparison of results.

A Series of Standardized Diagnostic Tests for the Fundamentals of Elementary Algebra. Harl D. Douglass. Journal of Educational Research, 1921, December, 396–403. A full description of a test of ten exercises for each of four fundamentals of elementary algebra. Emphasis is placed on accuracy.

Scale of Attainment No. 3.—For Measuring "Essential Achievement" in the Third Grade. Luella Pressey. Journal of Educational Research, 1921, December, 404–412. Description of a scale designed to measure the "fundamental promotion subjects" of Grade III; spelling, reading, and arithmetic. Results and first norms of the scale.

Comparing the Efficiency of Special Teaching Methods by Means of Standardized Tests. Samuel S. Brooks. Journal of Educational Research, 1921, December, 337–346. Seventh article by Superintendent Brooks on the general topic "Putting Standardized Tests to Practical Use in Rural Schools." Comparing new and old teaching methods in particular environments under controlled conditions.

A Threefold Experiment in High School English. R. H. Jordan. The English Journal, 1921, December, 560–569. Description of three tests: "(1) An attempt to determine the power of the students in interpreting and evaluating ordinary reading matter of contemporary interest and dignified style; (2) an attempt to determine the ability of pupils to classify verse according to merit; (3) a study of

the ability of the pupils to use the mechanics of the English language properly in simple composition." Discussion of results of the tests.

Intelligence Tests

Procedure Following a Testing Program. I. N. Madsen. School and Society, 1921, Dec. 24, 600–605. Tabulation of responses to a questionnaire sent to city schools in Idaho. Twenty-one school systems reported. Discussion of results of the various testing programs.

Porto Rico School Children and the Holley Picture Completion Test. Walter S. Monroe. School and Society, 1921, Dec. 24, 617–618. Comparison of test scores on the Holley Picture Completion Test of American and Porto Rico children. Close agreement.

Intelligence Tests and the Marks of Scholarship Men in College. J. A. Clement and W. E. Smythe. Educational Administration and Supervision, 1921, December, 510–516. An investigation of the intelligence test scores and marks of scholarship students in De Pauw University. Evidence shows the value of psychological tests in the selection of superior students.

Unreliability of Individual Scores in Mental Measurements. John L. Stenquist. Journal of Educational Research, 1921, December, 347–354. A plea for more thorough testing of individual pupils with data showing variations of performance in successive tests.

Miscellaneous

The Effect of Kinaesthetic Factors in the Development of Word Recognition in the Case of Non-readers. Grace M. Fernald and Helen Keller. Journal of Educational Research, 1921, December, 355–377. Seven case studies of children of normal mentality who could not read after three or more years in the public schools.

Backward Boys. Alice M. Clark. Pedagogical Seminary, 1921, December, 391–394. Illustrations of eminent men who were considered dull in boyhood.

A Study of 1000 *Errors in Latin Prose Composition.* C. W. Odell. School and Society, 1921, Dec. 31, 643–646. Classification of errors in Latin Prose and discussion of causes.

An Analysis of the Content of Six Third Grade Arithmetics. F. T. Spaulding. Journal of Educational Research, 1921, December, 413–423. Third Grade arithmetics show little standardization of amount of material covered and wide variation in the proportions of examples and problems presented. Favorable tendency toward problems dealing with human activities.

The Reliability of Prediction of Proportions on the Basis of Random Sampling. Ben Wood. Journal of Educational Research, 1921, December, 390–395. An experiment conducted sub-rosa by skeptics proves the reliability of predictions based on random sampling.

An Experiment Carried on with the Pupils of the Russell Pre-vocational Room. J. H. Vorhees. Journal of Educational Research, 1921, December, 378–389. Need for wider range of manual instruction for boys who are pedagogically retarded rather than the present emphasis on academic instruction. Eleven graphs show the present situation in Detroit with regard to such boys.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

1. *An Elementary Textbook in Educational Psychology.*—The author of this introduction to educational psychology[1] has in mind the beginning student in normal schools or colleges of education. It is uncommonly lucid in exposition, and comprehensive in general outline. It covers the topics of the traditional text of psychology and includes sections that represent a distinct innovation. The intention was to write a book "from the functional point of view though not leaning to behaviorism in its extreme form."

There is little likelihood that the author will be accused of "behaviorism in the extreme form" but it is quite likely that readers may feel that the author's point of view is not so thoroughly dynamic as the introduction would lead them to expect. The chapters on sensation, perception, memory and imagination, conception, thinking, attention, feeling and emotion and voluntary action are of the traditional sort, although ably and clearly written. The treatment of native activities and drives is relegated to a clearly subordinate position. The laws of associations are discussed, subordinate to the topics of mental imagery, in terms of the traditional recency, vividness, and primary distinctions. The treatment of learning is somewhat scanty.

Aside from the chapters on traditional topics there is one on language, one on individual differences and one on mental development; all of them good.

A distinct innovation is the inclusion of a chapter on each of the four main school subjects: reading, writing, spelling and arithmetic. These chapters are brief but remarkably suggestive introductions to the scientific work in these special fields. In an appendix samples of tests of general mental ability, achievement in school subjects, etc., are given.

As regards the underlying system, the book is conservative; there is little effort to develop new hypotheses, little special effort to make the facts conform to particular theories, little bias with reference to

[1] Cameron, E. H.: "Psychology and the School." New York: The Century Co., 1921, pp. XIV + 339.

any of the several schools of psychology. On the whole, the writings of Judd and Freeman have been drawn on with relative abundance. Because of its clear and concise treatment, this book will doubtless become widely used by those who desire a brief treatise on traditional psychology along with a conservative introduction to contemporary educational psychology.

A. I. G.

2. *A Study of Primary Children's Reading Interests.*[1]—The significance of the interest factor in primary reading materials has not been duly recognized by those who select and supply reading matter for use in the early grades. Before entering upon a discussion of her own investigation, Dr. Dunn summarizes previous studies along this line, showing how interest was inferred by various investigators. Evidence presented in a discussion and analysis of the content of twenty-nine school readers shows the lack of accepted criteria for inclusion and the need for careful studies along this line. The tabulations indicate that poetry constitutes fifty-one per cent of all selections. The actual relation between poetry and prose would perhaps be more truly set forth by using the page as a unit, in view of the fact that so many rhymes and verses are very brief. The same method would perhaps have the opposite effect when applied to fictitious stories, which, with their repetitions, are reported to make up forty-five per cent of the total number of selections. Of stories intended for belief, the investigator found only "a trace." While some primers were found to contain no formal drill material, others contained practically little else. There was further disagreement as to the grade placement of materials.

After a brief survey of the constitution of primary reading material during the past century, Dr. Dunn outlines her method of inquiry and presents the results. Thirty-one selections were submitted in pairs to one hundred and ninety-five classes. The children expressed their preference by ballot. A number of adult judges ranked the selections for twenty listed qualities. Because most samples are complex combinations of a number of interest factors, the method of partial correlations was used to eliminate each one of the nine most

[1] Dunn, Fannie Wyche, Ph. D.: "Interest Factors in Primary Reading Material. New York; Teachers College, Columbia University *Contributions to Education*, No. 113, 1921, p. 70.

significant factors from each of its fellows. This method of statistical analysis is carried into further detail, until the reader wonders whether the unanalytical and spontaneous story choices of children may be legitimately analyzed according to a scheme of qualities of which the adult judges alone were aware, and whether this is not a doubtful basis for the statistical method employed. The data points to the conclusion that children do not care for what adults consider humour and that verse form makes no definite appeal to children. "Adultness, style and other-sexness seemed to repel rather than attract." The leading positive interest factors for children in general are surprise and plot. "Animalness is raised to a level with surprise for boys and conversation shown to be of minor positive value for girls."

The study points to the need for further investigation and appraisal. "The neglected fields of fact . . . all need development. Not a crude rehearsal of ill-selected fact, but skillful composition incorporating salient interest-producing elements. There are few such books within the reading ability of primary children but there need be many more, opening doors into many fields. No field to which a dawning interest points should be excluded."

L. Z.

---

3. *Tests*[1] *of Trade Proficiency and Their Adaptation to Educational Use.*—In order to settle debatable issues with reference to vocational schools, means for measuring the human product of such institutions must be used. Adaptations of the army trade tests may solve the problem by facilitating comparison of students with men in the trades. Various methods of testing are compared and critically evaluated with reference to possible educational applications and a discussion of the technique of trade test construction takes up the possibility of predicting probable success or trade capacity and traces the development of prognostic measures. The methods used in an experiment in vocational guidance at the Manhattan Trade School for Girls, New York City, are discussed in this connection.

The final chapter gives observations on the narrowing effect of industrial training, the over-estimation of trade skill and intelligence

[1] Toops, Herbert Anderson, Ph. D.: Trade Tests in Education. New York: Teachers College, Columbia University *Contributions to Education*, No. 115, 1921, pp. VI + 118.

possessed by the typical journeyman and the relation of general education to proficiency in a trade. The use of self-administrative performance tests in trades is recommended as a means of providing incentive to learn. The investigator cites their use in two engineering schools. The obvious advantages of the plan are that the person tested can see how his performance compares with the norm or with any point on a scale without the loss of interest due to delayed scoring. The one-word-answer form of test is also self-scorable and this feature has been found to have a big appeal to the interest of students. These methods should appeal to the teacher because of their manifest economy in scoring time, and the fact that the pupil is convinced that his score is free from the effect of personal bias.

In the appendix there appears a chart for finding probable errors of Pearson r's, and a selected list of fifty references.

L. Z.

---

4. Professor Scott has written a plea for the recognition by employers of the individuality of their employees. The book contributes little to our knowledge of educational psychology, but the authors would probably be last to insist on making this a research contribution.

Mental tests in industry serve their greatest usefulness, not as a method of selection and elimination, but as a means of classification and adjustment of intelligence to difficulty of job.

Among office employees, distinctly higher average mental alertness scores for men than women indicate an occupational selection of the more capable men and less capable women in office work rather than a basic sex difference.

In simple, routine jobs the questionnaire method reveals that those men who are most badly retarded in school have the least "amount of dissatisfaction;" while, in jobs requiring high intelligence, those most retarded in school have the most dissatisfaction. These results—valuable if they can be substantiated—may be complicated by the questionnaire fallacy. The amount of dissatisfaction which one has for his work cannot be measured reliably by one or two questions. At best, the results will vary with the way in which the question is framed.

Some may be offended by the plain speech quoted to illustrate

---

[1] Scott, Walter Dill, and Hayes, M. H. S.: Science and Common Sense in Working with Men. New York: The Ronald Press, 1921, p. 154.

psychological principles. The pure psychologist may complain that the term "instinct" has neither been defined nor used according to his liking. The applied psychologist may wonder at the omission of charts, diagrams and statistics in a book dealing with tests and their applications; yet, were the book written otherwise, its message to employers, individuality of human beings and the great variability of human traits, might not be as widely distributed to those untrained in statistical methods. A realization by employers of the importance of individual differences, the authors feel, will do much to secure those industrial adjustments which exact measurement and many statistics may fail to bring about. Why should not someone now write a test primer to carry to the employee the message of individuality,—that message which has had such far-reaching beneficial effects in education?

HERBERT A. TOOPS.
Institute of Educational Research, Teachers College.

---

5. *The Growth of Intelligence.*—This problem is much disputed at the present time and the appearance of a monograph[1] on the subject is, therefore, very timely. The usual assumption has been that the rate of gain decreases gradually up to some age between fourteen and sixteen, at which point growth seems to stop. The author of this monograph, however, finds that the rate of growth is constant from nine to fifteen, and that there is no indication of cessation at this age. Other data, which he describes, suggest to him that growth continues up to eighteen at least. These conclusions are based upon three annual re-tests of 171 children, using a battery of eighteen tests. In addition the results of other workers are made use of to support his conclusions. The actual curves from his own data show a slightly decreasing rate of gain for memory functions, complex functions, and informational functions, but a constant rate of gain for simple functions. The decrease in the first three groups is assumed to be due to the selection of the cases tested, and, therefore, a constant rate is supposed to be truer to the facts. The author does not say why this assumption should not also apply to the simpler functions, in which case we should have an increasing rate of gain from age nine to age fifteen. The correlation between mental traits measured at a two-year interval is found to be high, thus strengthening our belief in the constancy of

[1] Brooks, F. D.: Changes in Mental Traits with Age. Teachers College, Columbia University *Contributions to Education*, No. 116, 1921.

the IQ and "indicating that these abilities are a relatively permanent endowment." The author finds no evidence of adolescent or pre-adolescent spurt in development as suggested by other workers, although he points out that irregularities in mental growth may occur in individual cases. Such irregularities can only be discovered by repeated tests of the same individuals over a long period of time. The monograph contains a very good digest of previous experimental work relating to this topic. An abstract of the monograph by the author himself has already appeared in the December, 1921, issue of this Journal.

R. PINTNER.

---

6. *A Careful Comparison of Achievement in Rural and City Schools.*[1] The rural schools have suffered by comparison with city schools in a number of recent investigations. This investigator contends that comparisons on the grade basis do not take account of some factors which are pertinent in determining the comparative efficiency of city and country schools. He cites great differences in length of school year and the differences in grade standards due to varying conditions under which instruction is carried on, and also calls attention to the necessity for taking retardation into account. He proposes to measure the progress of *all* pupils in terms of some unit common to all schools and suggests that progress made by pupils between the ages of seven and twelve, or ten and thirteen is a definite, universal measure. Standard Tests were used and the rural schools of Madison County, Kentucky, were compared with other town and country systems. The results seem to indicate that the six month schools are far less efficient than the nine month schools and that the latter compare very favorably with city schools.

L. Z.

II. Brief Editorial Notices of Mental and Educational Tests Recently Published

1. ENGEL, ANNA M. *Detroit First-grade Intelligence Test.* World Book Company, 1921.

A group examination to test general intelligence and to aid in the proper classification of children entering the first grade. The Ex-

[1] Frost, Norman: A Comparative Study of Achievement in Country and Town Schools. New York: Teachers College, Columbia University *Contributions to Education*, No. 111, 1921, p. 70.

aminer's Guide contains complete directions for administering and scoring the tests. The examination may be administered in from twenty to thirty minutes. Norms have been determined on the basis of 5000 Detroit pupils. Specimen set, containing 1 Examination, 1 Guide and 1 Record Sheet, 15 cts. postpaid. 25 Examination Booklets, including 2 Record Sheets, $1.50 net.

2. Miller, W. S. *Miller Mental Ability Test.* World Book Company, 1921.

A group intelligence test for grades 7 to 12 and for college freshmen. Time required—thirty to forty minutes. Norms have been established on the basis of 6000 high school pupils. Specimen set containing 1 Examination, 1 Key, 1 Manual, and 1 Age-Grade Score Sheet, 30 cts. net. Package of 25 examination booklets with Key, $1.00 net.

3. Downey, June E. *Downey Individual Will-temperament Test.* World Book Company, 1921.

A series of tests for determining the temperamental traits of an individual through motor reactions. No apparatus is required. No time limit. Norms are available. Specimen set containing 1 Test, 1 Record Card and a Manual, 20 cts. postpaid. Package of 25 Examination Booklets $1.00 net.

4. Ream, M. J. *Group Will-temperament Test.* Bureau of Personnel Research, Carnegie Institute of Technology, Pittsburgh, Pennsylvania, 1921.

A modification of the Downey Scale for use with groups of subjects. Time required—thirty minutes. No apparatus required.

5. Bureau of Educational Measurements and Standards, Kansas State Normal School, Emporia, Kansas, 1921.

Price List and Circular of Information on all intelligence and achievement tests, published by the World Book Company and distributed by this Bureau.

6. World Book Company. *Bibliography of Tests for Use in Schools.* 1921.

A booklet containing the titles, authors, and publishers of 294 intelligence and educational tests. Price 10 cts.

III. Announcement Concerning the Psychological Index

The Index for the Year 1920, No. 27, is now available, Psychological Review Company, Princeton, N. J. This is the *Annual Bibliography of the Literature of Psychology and Cognate Subjects* for the year 1920. Edited by Madison Bentley and Coleman R. Griffith.

IV. Brief Editorial Notices of Recent Educational Publications

1. Alexander, Carter and Theisen, W. W. *Publicity Campaigns for Better School Support.* (School Efficiency Monographs.) World Book Company, Yonkers, N. Y., 1921, pp. VII + 164. Paper.

A handbook for school superintendents and boards of education which gives the technique used in many successful school campaigns for increased support. Describes how to organize a campaign staff, who is to be reached by the campaign, how to organize the general campaign, how to select arguments and examples and to prepare and circulate material for effective publicity. It abounds in examples of good publicity material. Includes an excellent bibliography.

2. *An American Citizenship Course in United States History. Book I*, pp. X + 247. *Book II*, pp. X + 170. *Book III*, pp. X + 178. *Book IV*, pp. X + 251. *General Course*, pp. VI + 167. Published for the American Citizenship League. Chas. Scribners' Sons, New York, 1921.

3. Bruce, H. Addington. *Self-development.* New York: Funk & Wagnalls Co., 1921, pp. X + 332.

A series of essays of the "inspirational" type "for the ambitious." The usual prescriptions for weak will, memory, imagination, etc. are presented with an abundance of observations and illustrations but with an almost complete disregard of scientific knowledge.

4. Hertzog. W. S. *State Maintenance for Teachers in Training.* Baltimore: Warwick & York, 1921, pp. 144.

This monograph establishes the prevalence of a transient, generally

incompetent and immature body of teachers in American rural schools, a condition not nearly so true of city schools. Reports by a survey of conditions; the characteristics of our rural teaching population, the national situation with reference to teacher shortage, the development of colleges and other training agencies, 1870–1918. As a result of this evidence proposes plans for recruiting the profession through state subsidies and gives supporting data together with arguments from methods of recruiting other professions.

5. Jordan. R. H. *Nationality and School Progress, A Study in Americanization.* Bloomington, Illinois: Public School Publishing Company, 1921, pp. 105.

One of the *School and Home Monographs.* Reports an investigation of the records of school population in St. Paul and Minneapolis (shown to be typical of other cosmopolitan cities) to establish the relation between *progress in the schools* and nationality (in the first, second and third generations), as shown by relation to retardation, to acceleration, to school marks and to ability as determined by objective mental tests. Influence of nationality factor is shown in relation to mobility of students, occupation of parents, persistence of language in the home, economic status of parents, home conditions, church attendance and the like.

6. Pittman, Marvin S. *The Value of School Supervision.* Baltimore: Warwick & York, 1921, pp. X + 129.

Reports the author's pioneer experiment with a "Zone Plan" of supervision of rural schools proved to give results markedly superior to those obtained in other representative, but unsupervised, schools. Gives scientific evidence for the value of supervision and submits a tried and practicable plan. Results shown by (1) gains in abilities in school activities; (2) professional study of the teachers; (3) attendance, (4) effect on elimination of pupils from school; (5) social life of the community. We regard this monograph as of distinct value to district, county, and state school superintendents.

7. Webb, H. A. *General Science Instruction in the Grades.* Nashville, Tenn.: George Peabody College for Teachers, Contributions to Education, No. 4, 1921, pp. 105.

A quantitative analysis of eighteen textbooks in General Science,

leading to a statement of the current status of the following aspect of the field: (1) the subject matter of general science; (2) the acceptability of general science topics; (3) size of topics; (4) distribution of the sciences; (5) correlations between the sciences, (6) adaptability of general science; (7) analysis of marks; (8) analysis of complete reaction of children to science, etc.

H. O. R.

8. Myers, Caroline E. and Myers, Garry, C. *Measuring Minds.* New York: Newson & Co., 1921, pp. 55.

An examiner's manual to accompany the Myers Mental Measure. Contains directions, tables of norms, etc.

Brief Notices on New Books

9. Bridges, James Winfred. *An Outline of Abnormal Psychology.* Columbus: R. G. Adams & Co., 1921, pp. 226.

A second and revised edition of a manual which forms a very useful skeleton for a course on abnormal psychology. Contains very brief summaries of facts and theories, together with excellent lists of references.

V. Additional Publications Received

(*A*) *Books in General and Applied Psychology*

1. Anderson, J. B. *Applied Religious Psychology.* Boston: R. G. Badger Co., 1921, pp. 82.
2. Goddard, Henry H. *Juvenile Delinquency.*

(*B*) *Publications in the General Educational Field*

1. Bement, Alon. *Figure Construction.* New York: Gregg Publishing Co., 1921, pp. XII + 124. $2.50.
2. Berry, R. E. *An Analysis of Clerical Positions for Juniors in Railway Transportation.* (Part-time Education Series No. 6) Bulletin No. 5, Berkeley, California; University of California, 1921, pp. 104.
3. Betts, G. H. *The New Program of Religious Education.* New York: Abingdon Press, 1921, pp. 118. $0.75.
4. Claxton, P. P. and McGinniss, James. *Effective English, Junior.* Boston: Allyn & Bacon, 1921, pp. XV + 294.

5. Dadisman, Samuel H. *Methods of Teaching Vocational Agriculture in Secondary Schools.* Boston: Richard G. Badger, 1921, pp. 142. $2.00.
6. Ensign, Forest Chester. *School Attendance and Child Labor.* Iowa City Iowa: Athens Press, 1921, pp. IX + 263.
7. General Education Board, New York City.
   (1) *Public Education in Kentucky.* A state survey made by specialists of the General Education Board, 1921, pp. 213.
   (2) *Public Education in North Carolina.* A state survey made by specialists of the General Education Board. 1921, pp. 137.
8. Hayes, Augustus W. *Rural Community Organization.* Chicago: University of Chicago Press, 1921, pp. XI + 128. $1.50.
9. Hopkins, L. Thomas. *The Marking System of the College Entrance Examination Board.* (Harvard Monographs in Education, Series I, No. 2). Cambridge, Massachusetts: Graduate School of Education, Harvard University, 1921, pp. 15.
10. Lee, Jean Hunt, Johnson, Buford J. and Lincoln, Edith M. *Health Education and the Nutrition Class.* New York: E. P. Dutton & Co., 1921, pp. XV + 281. $3.50.
11. Lewis, E. E. *Scales for Measuring Types of English Composition.* Yonkers-on-Hudson, New York: World Book Company, 1921, pp. V + 142. $1.20.
12. North, S. M. *The Teaching of High School History.* State Department of Maryland, Baltimore, 1921, pp. 122.
13. *Old-age Support of Women Teachers.* "Studies in Economic Relations of Women," Vol. XI. Boston: Women's Educational and Industrial Union, 1921, pp. 122. $1.25.
14. O'Toole, Rose M. *Practical English for New Americans.* Boston: D. C. Heath & Co., 1921, pp. VI + 189.
15. Pryor, H. C. and Pittman, M. S. *A Guide to the Teaching of Spelling.* New York: MacMillan, 1921, pp. XI + 141.
16. Rice, O. S. *Lessons on the Use of Books and Libraries.* Chicago: Rand McNally & Co., 1920, pp. XVIII + 178.
17. Smith, J. Russell. *Human Geography. Book I.* Chicago: John C. Winston Co., 1921, pp. VI + 369.
18. *Spelling Survey in the Schools of Newark, New Jersey.* Newark: Board of Education, 1920, pp. 32.
19. Tigert, John James. *The Teaching of Civics.* Washington: Bureau of Education, 1921, pp. 10.
20. *Virginia Public Schools, Part II: Educational Tests.* Yonkers-on-

Hudson, New York: World Book Co., 1921, pp. XII + 235. $2.40.

C. New School Textbooks

1. Dance, E. H. *A Constructive Note-book of English History.* Oxford, England: Basil Blackwell, 1921, pp. XII + 110.
2. Bedford, Edgar A. *General Science.* Boston: Allyn & Bacon, 1921, pp. XXIV + 387.
3. Dunn, Arthur W. *Community Civics for City Schools.* Boston: D. C. Heath & Co., 1921, pp. X + 582.
4. Ermeling, Willard W.; Fisher, Ferdinand A. P. and Greene, George G. *Mechanical Drawing.* Milwaukee: Bruce Publishing Co., 1921, pp. 78. $0.45.
5. Flagg, Mildred B. *Community English.* New York: MacMillan, 1921, pp. XVI + 266.
6. Herben, Beatrice Slayten. *Jack O'Health and Peggy O'Joy.* New York: Charles Scribner's Sons, 1921, pp. IV + 39.
7. Huff, Bessie M. *A Laboratory Manual for Journalism in High Schools.* Muskogee, Oklahoma: Star Printery, 1921, pp. VIII + 97.
8. Jackson, Bennett B.; Deming, Norma H. and Bemis, Katherine I. *Opportunities of Today for Boys and Girls.* New York: Century Co., 1921, pp. XII + 274.
9. Klenke, William W. *Art and Education in Wood Turning.* Peoria, Illinois: Manual Arts Press, 1921, pp. 100. $1.40.
10. McMurry, F. M. and Parkins, A. E. *Elementary Geography.* New York: MacMillan Co., 1921, *Part I,* pp. V + 175; *Part II,* pp. V + 157.
11. Myers, P. V. N. *General History.* Boston: Ginn & Co., 1921 (revised), pp. XIV + 711 + XXXIII. $2.00.
12. Potter, Milton C.; Jeschke, H. and Gillet, Harry O. *Oral and Written English, Intermediate Book.* Boston: Ginn & Co., 1921, pp. VIII + 270 + XXIII.
13. *Standard Library Organization and Equipment for Secondary Schools.* (New York State Library School Bulletin) No. 45. Albany: University of the State of New York, 1920, pp. 39.
14. Starch, David and Myrick, George A. *The Test and Study Speller.* Silver Burdette & Co., Boston, 1921. Three Books.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Vol. XIII March, 1922 No. 3

THE CORRELATIONS OF ACHIEVEMENT IN SCHOOL SUBJECTS WITH INTELLIGENCE TESTS AND OTHER VARIABLES[1]

ARTHUR I. GATES

Teachers College, Columbia University

Many tests of intelligence are now available. They are classified in several ways; most commonly as verbal and non-verbal, although many combine both varieties of exercises. Whether the extremes have equal predictive value is but imperfectly known. In one sense they have equal predictive value, since perhaps every function is closely associated with certain other functions. The question is: Just what functions are correlated highly with verbal, non-verbal or any other abilities?

It has often been demonstrated that certain tests have a high reliability; that is, they measure consistently whatever it is that they do measure. The Stanford-Binet measures something very well but of just what abilities it is a valid measure or for just what abilities it gives a valid prediction, is far from perfectly known.

Usually the test has been developed to yield a predictive measure of some particular ability or composite of abilities. Often the tests reach a high state of mechanical perfection, while the criterion—the abilities to be predicted—remains unanalyzed and but imperfectly measurable. In the field of educational prediction, the crucial problem is the perfection of criteria. We secure little or no information about a test by computing correlations with teachers' estimates, which may be good or bad—mediocre, in all likelihood—but just which is never known. If the actual correlation of the Stanford-Binet and real "general intelligence" were approximated unity, we would not dis-

[1] A contribution of the Department of Educational Research of the Scarborough School, Scarborough, N. Y.

cover it by comparisons with the conventional school grades. We need better criteria.

In this study, we have taken as our problem the prediction of achievement in the fundamental school subjects. A special effort has been made to secure measures of achievement which are as valid and reliable as facilities would permit. Each school subject has been measured by a large number of standard tests: In some cases, *e.g.*, reading in certain grades, nearly 10 hours work is represented. The reliability of the composites is high; the coefficients of reliability (self correlations) being in all cases between 0.92 and 0.97.

We cannot be so sure of the validity of the composites. The series of arithmetic tests, even if self consistent, may be an inadequate measure of "general arithmetical ability." Without an ultimate criterion outside the tests themselves, the validity of the tests cannot be determined. The different tests should be weighted by use of the regression coefficients, but with no ultimate criterion the perfect weights cannot be determined. Mere accumulation of inadequate tests or of tests which measure well only a common fraction of the "general" ability will not, of course, insure a valid criterion. The best one can do is to select the tests which are most reliable, which test as diverse elements within the general function as possible, and weight them arbitrarily.

Even with reliable and valid tests, properly weighted, as a criterion, perfection is not necessarily attained. The effects of school and home emphasis on particular subjects may operate to reduce the correlation with the predictive variables. Arithmetic may be so stressed that the pupils have attained nearly maximal performance ability but spelling may be so neglected that few if any have it, even approximately. In a second school, conditions may be the reverse. The predictive value of the intelligence tests will almost certainly differ in the two schools. The ideal would be a situation in which each pupil had reached his approximate limit of improvement which is then measured by valid and consistent instruments. This situation will probably not be obtained except under strictly experimental conditions.

The Tests Employed

1. *Tests of Achievement:*

The measures for grades I and II are very much less extensive and valid than those in the other grades. The tests given were:

(*A*) For grade I. Indiana Reading vocabulary, Haggerty, Sigma I, and 50 words from Ayres Spelling list.

(*B*) For grade II. Indiana Reading vocabulary, Haggerty, Sigma I, Thorndike-McCall Reading, 50 words from Ayres Spelling and the Indiana Composite Scale of Attainment, No. 2, including tests of reading, word knowledge, arithmetic, and spelling.

The following tests were given in grades III to VIII inclusive:

(*A*) *Composite of Comprehension in Reading.*—The scores of the Burgess, Courtis, Monroe, Thorndike, Thorndike-McCall, and Woodworth-Wells Directions tests were combined. The scores for certain tests were the means of several tests in which different forms or editions of the test were used. In grades IV and VI the Thorndike-McCall was given five times. The SD's were such that the several tests were given approximately equal weights. This is true of the following composites.

(*B*) *Rate of Reading.*—A composite of the Burgess, Courtis Rate, Brown Rate and Monroe Rate, all given at least twice and weighted equally.

(*C*) *Arithmetic.*—A composite of the Woody 30-minute test for each of the four operations, the Monroe Diagnostic, 11 to 21 specific tests, ranging from easy integer combinations to fractions and decimals, and the Monroe Reasoning Test.

(*D*) *Spelling.*—Four tests making a total of 186 words of varied difficulty from the Ayres-Buckingham list.

(*E*) *Writing.*—Two sets of specimens graded for quality by the Thorndike scale. Speed and quality were combined by multiplying the number of letters per minute by the quality score.

(*F*) *Composite of General Achievement.*—A combination of reading comprehension, reading rate, arithmetic and spelling. The SD's were such that each received approximately the same weight.

2. *Other Criteria Used:*

(*A*) *Teachers' Estimates of School Attitudes.*—A rating scale requesting judgments on a composite of industry, interest, attention, etc. was filled out for each pupil by from 5 to 9 teachers or supervisors, independently. Only the ratings for grades IV to VIII are used in this study.

(*B*) *Chronological age.*

3. *The Predictive Measures:*

(*A*) *Stanford-Binet Mental Age*, grades I to VI, inclusive.

(*B*) *Group Tests of Intelligence.* (1) Dearborn, Examinations 1, 2,

3, and Total, (2) Dearborn, Examinations 4, 5, and Total, (3) Haggerty, Delta 1, (4) Haggerty, Delta 2, (5) Holley's Picture Completion, (6) Holley's Sentence Vocabulary, (7) Illinois Examination, (8) Kingsbury Primary, Form A, (9) Meyers Mental Measure, (10) National Intelligence, Form A, B, and Total, (11) Otis, Primary, Form A, (12) Otis, Advanced, (13) Pressey Primer, (14) Terman Group, (15) Thorndike-McCall Reading Test.

These tests were given as follows: To grades I and II, Nos. 1, 3, 5, 8, 9, 11 and 13; grade III, Nos. 1, 3, 6, 7, 9, 10, 11 and 15; grades IV 2, 4, 5, 6, 7, 9, 10, 11 and 15; grade V to VIII inclusive, 2, 4, 5, 6, 7, 9, 10, 12, and 15. The Terman Group Test, grades VII and VIII.

The Subjects.—The subjects were the pupils of grades I to VIII inclusive, in the Scarborough School, at Scarborough, N. Y. These are select groups; the mean Stanford-Binet Intelligence Quotients are about 117. There were about 20 pupils to the grade.

Statistical Methods Employed.—All coefficients of correlation were computed by the Pearson Product-Moment formula. The tables give the correlations for each variable for each grade separately. The mathematical estimates of the probable error of the correlations which are based on the number of cases are not included, but may be readily estimated. The variability of the separate grade (or test) correlations from the mean of the grades (or tests) gives the best notion of their reliability. The mean is used as a measure of central tendency of the r's and the SD's give the variability.

Because of the restriction in the range of abilities in our groups (children of like abilities being selected for a grade) the correlations are certainly lower than they would be for an unselected group. There is no technique for correcting the attenuation due to restriction in range, which could be applied to our data.

In so far as the measures used fail to measure a subject's abilities perfectly, the correlations will again be attenuated. Many of our measures have a high degree of reliability because of the thoroughness of the testing and since our main purpose is make comparisons of one variable with another, within the data for our own groups, we have made no correction for this type of attenuation. It will, of course, be understood that the absolute amounts of our coefficients may not be compared with those obtained from other groups unless the range should happen to be the same.

In many cases, the inter-correlations are so many and varied that they cannot be readily interpreted by inspection. It has been advis-

able to employ, in certain cases, the technique of multiple and partial correlations, when the most perfect weight of each variable has been determined by the regression equations.

There are now several methods by which these data may be obtained. The most recent and excellent procedure in print is that devised by Dr. T. L. Kelly.[1] Dr. Herbert A. Toops has more recently devised (but not yet printed) a different and more expeditious method of obtaining multiple correlations which will be explained and defended by him in due time. All of the partial correlations presented in this article were computed by Dr. Kelley's method and the multiple correlations by Dr. Toops' method. In certain cases multiple correlations and partial correlations of the first order were computed by both methods and found to agree.

Part I. Results of Tests in Grades I and II

Table I gives the inter-correlations of the group tests, and the correlations of the group tests with Stanford Mental Age and with the composite of tests of achievement for each grade, together with the average results for the two grades.

1. *Correlations with the Composite of Achievement.*—Averaging the results of all tests for the two grades the following correlations are found:

1. Achievement with MA.................... 0.40
2. Achievement with Group Tests[2].............. 0.27
3. Achievement with Chronological Age......... 0.15
4. MA with Group Tests..................... 0.43
5. MA with Chronological Age.................. 0.27
6. Group Tests with Chronological Age.......... 0.47

The inter-correlations are too complicated to permit ready interpretation by inspection. While MA yields a higher correlation with achievement than any single test, group tests and age also show positive correlations. Do the latter criteria add anything not included in the MA? If so what and how much?

---

[1] Kelly, T. L.: Chart to Facilitate the Calculation of Partial Coefficients of Correlation and the Regression Equations. Stanford University Publication, School of Education, *Monograph* No. 1, 1921.

[2] This is not the correlation with a composite of group tests; it is the correlation of the average (single) group test when the r's for the two grades are averaged.

TABLE I.—CORRELATIONS FOR GRADES I AND II

Grade	1 Dearborn 1	2 Dearborn 2	3 Dearborn 3	4 Dearborn total	5 Otis Primary	6 Kingsbury	7 Haggerty Delta 2	8 Pressey Primer	9 Holley Picture	10 Myers	11 Chronological Age	12 Stanford MA	13 Composite Achievement	14 Mean Group Tests	15 Mean all criteria	16 Time (minutes)
Correlations of Dearborn, Examination 1 with																
I	....	0.84	0.67	0.91	0.82	0.53	0.55	0.34	0.45	0.62	0.48	0.53	0.42			
II	....	0.43	0.75	0.82	0.56	0.14	−0.13	0.43	0.18	0.51	0.42	0.54	0.10			
Mean	....	0.64	0.71	0.87	0.69	0.34	0.21	0.39	0.31	0.57	0.45	0.54	0.26	0.48	0.47	40
Correlations of Dearborn, Examination 2, with																
I	0.84	....	0.69	0.95	0.71	0.46	0.53	0.21	0.32	0.47	0.44	0.53	0.53			
II	0.43	....	0.36	0.61	0.56	0.25	0.49	0.31	0.25	0.41	0.40	0.32	0.38			
Mean	0.64	....	0.53	0.78	0.64	0.36	0.51	0.26	0.29	0.44	0.42	0.43	0.46	0.46	0.46	40
Correlations of Dearborn, Examination 3, with																
I	0.67	0.69	....	0.81	0.70	0.52	0.50	0.51	0.60	0.73	0.43	0.52	0.42			
II	0.75	0.36	....	0.81	0.67	0.18	0.17	0.51	0.20	0.49	0.49	0.58	0.12			
Mean	0.71	0.53	....	0.81	0.69	0.35	0.34	0.51	0.40	0.61	0.46	0.55	0.27	0.52	0.50	40
Correlations of Dearborn, Total, with																
I	0.91	0.95	0.81	....	0.76	0.52	0.63	0.35	0.49	0.63	0.53	0.58	0.48			
II	0.82	0.61	0.81	....	0.69	0.30	0.24	0.59	0.25	0.68	0.45	0.56	0.32			
Mean	0.87	0.78	0.81	....	0.73	0.41	0.44	0.47	0.37	0.66	0.49	0.57	0.40	0.51	0.51	120
Correlations of Otis Primary with																
I	0.82	0.71	0.70	0.76	....	0.45	0.53	0.36	0.30	0.68	0.40	0.46	0.38			
II	0.56	0.56	0.67	0.69	....	0.58	0.61	0.54	0.38	0.75	0.38	0.55	0.20			
Mean	0.69	0.64	0.69	0.73	....	0.52	0.57	0.45	0.34	0.72	0.39	0.51	0.29	0.60	0.56	22
Correlations of Kingsbury with																
I	0.53	0.46	0.52	0.52	0.45	....	0.24	0.40	0.80	0.47	0.54	0.56	0.10			
II	0.14	0.25	0.18	0.30	0.58	....	0.26	0.32	0.58	0.51	0.54	0.15	0.08			
Mean	0.34	0.36	0.35	0.41	0.52	....	0.25	0.36	0.69	0.49	0.54	0.36	0.09	0.32	0.38	7
Correlations of Haggerty, Delta I with																
I	0.55	0.53	0.50	0.63	0.53	0.24	....	0.38	0.39	0.45	0.51	0.36	0.21			
II	−0.13	0.49	0.17	0.24	0.61	0.26	....	0.39	−0.03	0.31	0.31	0.44	0.60			
Mean	0.21	0.51	0.34	0.44	0.57	0.25	....	0.39	0.18	0.38	0.41	0.40	0.41	0.36	0.37	12
Correlations of Pressey Primer with																
I	0.34	0.21	0.51	0.35	0.36	0.40	0.38	....	0.54	0.59	0.76	0.24	0.17			
II	0.43	0.31	0.51	0.59	0.54	0.32	0.39	....	0.26	0.38	0.30	0.52	0.26			
Mean	0.39	0.26	0.51	0.47	0.45	0.36	0.39	....	0.40	0.49	0.53	0.38	0.22	0.41	0.39	12
Correlations of Holley Picture Completion with																
I	0.45	0.32	0.66	0.49	0.30	0.80	0.39	0.54	....	0.56	0.38	0.35	0.14			
II	0.18	0.25	0.20	0.25	0.38	0.58	−0.03	0.26	....	0.50	0.40	0.10	0.02			
Mean	0.31	0.29	0.43	0.37	0.34	0.69	0.18	0.40	....	0.53	0.39	0.23	0.08	0.39	0.39	5
Correlations of Myers Mental Measure with																
I	0.62	0.47	0.73	0.63	0.68	0.47	0.45	0.59	0.56	....	0.58	0.16	0.13			
II	0.51	0.41	0.49	0.68	0.75	0.51	0.31	0.38	0.50	....	0.51	0.41	0.25			
Mean	0.57	0.44	0.61	0.66	0.72	0.49	0.38	0.49	0.53	....	0.55	0.29	0.19	0.43	0.49	10
Correlations of Stanford Mental Age with																
I	0.53	0.53	0.52	0.58	0.46	0.56	0.36	0.24	0.35	0.16	0.30	....	0.36			
II	0.54	0.32	0.58	0.56	0.55	0.15	0.44	0.52	0.10	0.41	0.24	....	0.44			
Mean	0.54	0.43	0.55	0.57	0.51	0.36	0.40	0.38	0.23	0.29	0.27	....	0.40			

The questions may be answered by the use of multiple and partial correlations. These have been computed by the formulæ earlier mentioned. The regression equation gives the weights that each variable should be given to predict achievement most perfectly. They are:

7. Weight of MA.............................. 1.00
8. Weight (β) of Group Tests.................. 0.345[1]
9. Weight (β') of Chronological Age............ 0.008[1]

Chronological Age thus appears to add little which is independent of MA and group tests, but the group tests add something to MA which is independent of it.

For practical purposes we wish to know whether the addition of the independent contributions of Group Tests and Chronological Age to MA will increase the correlation with achievement sufficiently to justify the time and expense of administering them. The multiple correlation, each variable weighted perfectly according to its independent addition, will give this information.

10. Simple r, Achievement with MA........... 0.40
11. Multiple r, Achievement with (MA + Group) 0.415
12. Multiple r, Achievement with (MA + Group + CA)................................ 0.416

If the Stanford-Binet is given, very little is added to the correlation value by adding a group test or CA, even when they are perfectly weighted. The MA alone is clearly superior to an average Group Test or Chronological Age alone (see 1, 2 and 3 above).

All of the Group Tests (with exception of parts of the Haggerty) are composed of non-verbal materials, so that the results just enumerated cannot be generalized to include verbal material. The relative predictive values of the MA and group tests will be taken up again in connection with the higher grades where verbal as well as non-verbal tests are available.

The validity of the correlations just considered depends upon the validity of the criterion, the composite of achievement. It is less valid than is desirable because of the dearth of adequate educational

[1] β is the weight of group tests independent of the elements already given in the correlation of MA with the criterion; β' is the weight of chronological age independent of the other two. If chronological age were put in second place instead of third it would show a larger weight than it does. It is customary to arrange the variables in the order of the magnitude of their simple correlation with the criterion.

tests for these grades and because difference in pre-school training may considerably affect achievement, particularly in grade I.

2. *The Relation between the Length of a Group Test and Its Correlation with Criteria.*—In Table I the inter-correlations of Group Tests are averaged in column 14. Column 15 gives the mean of all correlations except those with Chronological Age. Column 16 gives the approximate working time for each tests. The group tests are alike in being composed of non-verbal material with the exception of the Haggerty Delta I which contains some verbal material.

Inspection of columns 14, 15 and 16 shows a close relation between the length of the test and the magnitude of the correlations. The range for time, however, is large compared to the range of the correlations. If the ranges are made equal by the use of the rank method, the correlation of time with mean inter-correlations (column 14) is 0.75; of time with correlations for all criteria (column 15) is 0.69. Leaving the ranges as they are, using the Product-moment formula, the correlations become much smaller, *viz.* 0.50 and 0.49. In either case, the longer tests, in general, have a higher predictive value but our data are too few to permit an estimate of the increments in r, which are produced by given additions to the length of the test.

Part II. Correlations in Grade III

Grade III offers an opportunity to test the relative validity of verbal and non-verbal tests for the reason that there are available a number of tests constructed exclusively of each type of material. Our measures of achievement in school subjects are much more extensive and valid in this grade than in the primary grades. The tests used in constructing the composites for the school subjects are listed in the first section of this paper.

Table II gives the detailed results. Tests numbered on lines 1 to 7 inclusive are non-verbal. Line 8 gives the mean inter-correlations for this group. Tests on lines 10–15 inclusive are verbal, wholly or chiefly. Line 16 gives the mean inter-correlations for this group. In computing the means, the correlations of parts of a test with the whole (Dearborn and National) have been omitted since the correlation is made much higher by the correlation of the part with itself in the whole. The total working time for the groups of non-verbal tests is about one-third greater than that required for the group of verbal tests (see column 21).

1. *Correlations of Verbal and Non-verbal Tests with Themselves and Each with the Other.*

From Table II the following means were computed:

Mean inter-correlation of non-verbal tests	0.40
Mean inter-correlation of verbal tests	0.62
Mean correlation, non-verbal with verbal	0.24

The verbal tests quite clearly yield different results from the non-verbal. The verbal tests are more consistent with themselves than the non-verbal but the latter agree among themselves better than with the verbal. Comparative data on particular tests are available in Table II.

2. *The Independent Values of MA Verbal, and Non-verbal Group Tests for Predicting School Achievement.*

The following are the mean simple correlations for this grade:

1.	Achievement with verbal group	0.65
2.	Achievement with MA	0.47
3.	Achievement with non-verbal	0.22
4.	Verbal group with MA	0.47
5.	Verbal group with non-verbal	0.24
6.	MA with non-verbal group	0.16

The verbal group test appears to be the best single predictive measure, followed closely by MA. The correlation of MA and the verbal group is the same as MA with achievement. The other correlations are lower.

Does MA or non-verbal group really add anything unique to the verbal group for predicting achievement?

By use of the regression equation the following weights were found.

7.	Weight of verbal group	1.000
8.	Weight (β) of MA	0.499
9.	Weight (β') of non-verbal	0.097

Non-verbal adds very little not already accounted for by MA and the group tests, while MA shows a fair contribution, independent of the verbal tests.

The following multiple correlations show to what extent the addition of MA and non-verbal increases the correlation with achievement.

10.	Simple r, achievement with verbal	0.65
11.	Multiple r, achievement with (verbal + MA)	0.699
12.	Multiple r, achievement with (verbal + MA + non-verbal)	0.702

The properly weighted composite of the verbal group test and the MA yields a very high correlation with achievement. Non-verbal tests add very little to the combination. Whether the 0.05 increase over verbal produced by adding MA is worth while, is after all an

TABLE II.—CORRELATIONS FOR GRADE III[1]

	1 Dearborn 1	2 Dearborn 2	3 Dearborn 3	4 Dearborn total	5 Otis Primary	6 Haggerty	7 Meyers	8 Holley Vocabulary	9 National Intelligence Test A	10 National Intelligence Test B	11 National Intelligence Test Total	12 Illinois	13 Thorndike-McCall	14 Chronological Age	15 Stanford Mental Age	16 Reading Comprehension	17 Reading Rate	18 Arithmetic	19 Spelling	20 Composite Achievement	21 Approximate time (minutes)
Non-verbal tests																					
1. Dearborn 1	....	0.42	0.48	0.79	0.53	0.23	0.36	0.18	0.26	0.10	0.19	0.47	0.15	0.20	0.14	0.18	0.16	0.10	0.08	0.21	40
2. Dearborn 2	0.42	....	0.23	0.60	0.28	0.31	0.22	0.11	0.60	0.18	0.38	0.31	0.26	0.14	0.10	0.21	0.14	0.21	0.18	0.30	40
3. Dearborn 3	0.48	0.23	....	0.78	0.34	0.44	0.37	0.30	0.20	0.14	0.22	0.47	0.03	0.13	0.11	0.10	0.15	0.22	0.12	0.20	40
4. Dearborn total	0.79	0.60	0.78	....	0.52	0.59	0.45	0.26	0.41	0.19	0.33	0.48	0.24	0.20	0.15	0.20	0.18	0.18	0.13	0.30	120
5 Otis Primary	0.53	0.28	0.34	0.52	....	0.42	0.49	0.10	0.24	0.08	0.28	0.17	0.13	0.50	0.18	0.03	0.02	0.27	0.11	0.18	22
6. Haggerty	0.23	0.31	0.44	0.59	0.42	....	0.38	0.25	0.20	0.08	0.26	0.58	0.10	0.16	0.24	0.09	0.07	0.12	0.13	0.18	12
7. Meyers	0.36	0.22	0.37	0.45	0.49	0.38	......	0.27	0.19	−0.04	0.15	0.36	0.02	0.15	0.20	0.02	0.07	0.08	0.16	0.19	10
8. Mean	0.40	0.29	0.37	0.52	0.43	0.40	0.38	0.21	0.30	0.10	0.26	0.41	0.13	0.21	0.16	0.12	0.11	0.17	0.13	0.22	164[2]
9. SD	0.10	0.07	0.09	0.07	0.09	0.11	0.08	0.07	0.14	0.07	0.07	0.13	0.09	0.12	0.05	0.07	0.06	0.06	0.03	0.05	
Verbal tests																					
10. Holley Vocabulary	0.18	0.11	0.30	0.26	0.10	0.25	0.27	....	0.65	0.59	0.63	0.76	0.58	−0.26	0.33	0.69	0.64	0.23	0.58	0.64	15
11. National Intelligence Test A	0.26	0.60	0.20	0.41	0.24	0.20	0.19	0.65	....	0.63	0.79	0.63	0.57	0.08	0.44	0.77	0.65	0.47	0.63	0.68	17
12. National Intelligence Test B	0.10	0.18	0.14	0.19	0.08	0.08	−0.04	0.59	0.63	......	0.74	0.43	0.76	−0.33	0.38	0.67	0.70	0.38	0.64	0.60	16
13. National Intelligence Test A + B	0.19	0.38	0.22	0.33	0.28	0.26	0.15	0.63	0.79	0.74	....	0.61	0.80	−0.16	0.53	0.69	0.77	0.46	0.71	0.78	33
14. Illinois	0.47	0.31	0.47	0.48	0.17	0.58	0.36	0.76	0.63	0.43	0.61	....	0.38	−0.01	0.23	0.59	0.46	0.20	0.34	0.52	17
15. Thorndike-McCall	0.15	0.26	0.03	0.24	0.13	0.10	0.02	0.58	0.57	0.76	0.80	0.38	....	−0.16	0.16	0.88	0.83	0.38	0.45	0.70	30
16. Mean	0.23	0.31	0.23	0.32	0.17	0.25	0.16	0.64	0.62	0.60	0.68	0.56	0.62	−0.15	0.35	0.72	0.68	0.35	0.56	0.65	95[2]
17. SD	0.12	0.16	0.14	0.10	0.07	0.16	0.14	0.06	0.03	0.11	0.09	0.14	0.11	0.14	0.12	0.09	0.12	0.10	0.13	0.08	
18. MA	0.14	0.10	0.11	0.15	0.18	0.24	0.20	0.33	0.44	0.38	0.53	0.23	0.16	0.10	....	0.29	0.20	0.42	0.32	0.47	

[1] In computing the mean intercorrelations, the coefficients of part with whole and whole with part (National Dearborn) have been omitted.
[2] Totals, in which times for composite tests (National Intelligence Test and Dearborn) are included but once.

administration problem depending on the circumstances and needs of the school.

3. *The Relation of the Validity of the Test to Its Length.*—Fairly high correlations were found between the length of the non-verbal tests and the magnitude of the *r's* with the criterion in the case of the primary grades. In grade III, the correlations with composite of achievement will be used as a criterion for the reason that it is probably more reliable than any other. The correlations (Product Moment formula) of the length of the test (column 21, Table II) and the *r's* with achievement (column 20) are for non-verbal tests 0.76. The correlation between length of the verbal tests and the magnitude of the *r's* with achievement is 0.81. For both types of material, even with other things (methods of weighting in scoring, differences in validity of constituent exercises, etc.) varying as they do, the longer the test the higher the predictive value.

4. *Correlations with Achievement in Particular School Subjects.*—Columns 16 to 20 give the data, which are here summarized:

	Correlation of Mean Non-verbal	Correlation of Mean Verbal
1. With Reading Comprehension...	0.12	0.72
2. With Reading Rate............	0.11	0.68
3. With Spelling..................	0.13	0.56
4. With Arithmetic...............	0.17	0.30
5. With Composite of Achievement.	0.22	0.65

Since the measures of achievement are largely verbal the fact that the verbal intelligence tests yield much higher correlations with them need not be surprising. Arithmetic is commonly judged to be less verbal than reading or spelling and the verbal intelligence tests, in fact, yield the lowest correlations with arithmetic, but the correlation of non-verbal with arithmetic is still lower. None of our material gives a useful prediction of this function. Reading and spelling, being decidedly verbal, yield high correlations with verbal group tests.

5. *Correlations with Chronological Age.*—Column 14, Table II, yields the following summary:

Mean *r*. non-verbal with age $+0.21$
Mean *r*. verbal with age -0.14

The data are in substantial agreement with the facts found in the case of the primary grades. Why non-verbal tests tend, much more than verbal, to be positively associated with age in the case of school grades, cannot be objectively determined by our data.

(*To be concluded in April*)

INTELLIGENCE IRREGULARITY AS MEASURED BY SCATTERING IN THE BINET SCALE[1]

J. E. W. WALLIN

Bureau of Special Education

(Subnormal and Delinquent Children)

Miami University

I shall attempt here to give a brief resumé of a number of studies, of unevenness in intelligence development as measured by scattering in the Binet-Simon scale which I have made during the last decade. As here employed, *scattering* signifies the number of tests passed in the Binet-Simon scale above the basal age. Our analyses of scattering thus far published have been confined to the 1908 and 1911 Binet-Simon scale, and have been based on 333 epileptics in an institution,[2] on 34 psychotics in institutions,[3] and on 1181 consecutive subjects of various types coming to a university and a public school psycho-educational clinic, classified both according to intelligence age and according to diagnostic category (determined to a considerable extent by grade of intelligence), 840 of whom were thoroughly tested and 341 less thoroughly tested.[4] All except two of the insane were adults. Thirty per cent of the epileptics were under 21 years of age, the youngest being 5 years old.[5] In the group of 840 school cases the average chronological age of the boys was 11.10, of the girls 10.87, and of both sexes 11.07; while the average intelligence age by the 1908 scale was 8.59 for the boys, 8.10 for the girls, and 8.45 for both; and by the 1911 scale 8.16 for the boys, 7.43 for the girls, and 7.95 for both. For the

[1] Presented, in extract, before the Section of Clinical Psychology of the American Psychological Association, December 28, 1921.

[2] "Experimental Studies of Mental Defectives," 1912, p. 22—an incomplete analysis which cannot be completed because of the inaccessibility of the original records.

[3] "Problems of Subnormality," 160f (of 1921 reprint).

[4] The Phenomenon of Scattering in the Binet-Simon Scale. *Psychological Clinic*, 1917, pp. 179–195.

Wide Range Versus Narrow Range Binet-Simon Testing. *Journal of Delinquency*, 1917, pp. 315–330.

A Further Comparison of Scattering and of the Mental Rating by the 1908 and 1911 Binet-Simon Scales. *Journal of Delinquency*, 1918, pp. 12–27.

An Analysis of Binet-Simon Records. *School and Society*, March 30, 1918.

[5] The distribution of the chronological ages for the children is given in "Experimental Studies of Mental Defectives," 1912, p. 102.

group of 341 the average chronological age was 10.4 for the boys, 10.1 for the girls, and 10.3 for both sexes.

Since the publication of the above studies we have finished an analysis of scattering by the same method[1] among 1025 consecutive

TABLE I.—DISTRIBUTION OF CHRONOLOGICAL AND INTELLIGENCE AGES

Ages	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	28	Total
Chronological:																				
Number boys	..	1	..	2	4	34	96	158	132	109	84	68	27	8	6	1	1	2	1	734
Number girls	..	1	..	...	3	16	38	41	44	47	40	34	9	10	3	3	..	2	..	291
Both	..	2	..	2	7	50	134	199	176	156	124	102	36	18	9	4	1	4	1	1025
Intelligence:																				
Number boys	5	19	32	65	156	162	153	85	32	17	3	1	0	1	..	..	..	..	..	731
Number girls	5	13	20	39	65	69	56	11	4	2	0	0	0	0	..	..	..	..	..	284
Both	10	32	52	104	221	231	209	96	36	19	3	1	0	1	..	..	..	..	..	1015[1]

Average chronological ages: boys 10.47; girls 10.96; both 10.61. The average intelligence ages are: boys 7.01; girls 6.71; both 6.92.

A few cases have been omitted in this series because an exact intelligence age could not be assigned or because the records were inadvertently not transferred.

cases examined by means of the Stanford-Binet in a school psycho-educational clinic, when classified both according to intelligence age and diagnostic category. The distribution of the chronological and intelligence (Binet-Simon) ages for this group is shown in Table I.

The average chronological ages in the three groups of public school clinic cases do not differ significantly. But the average intelligence age is perceptibly lower in the Stanford, amounting to one year as compared with the 1911 scale, and about a year and a half as compared with the 1908 scale. The lower rating in the Stanford is probably due to the fact that this scale is more difficult, as shown by earlier analyses.[2]

THE RELATION OF SCATTERING TO GRADE OF INTELLIGENCE

Binet and Simon[3] asserted that the "defective child" is "inferior, not in degree, but in kind. The retardation of his development has not been uniform. . . . So far as certain faculties are concerned, he remains on the level of a younger child; but in respect to others, he is on a level with normal children of his own age. An unequal and imperfect development is consequently his specific characteristic.

[1] The detailed data will be published elsewhere.

[2] *E.g.*, "The Results of Retests by the Binet Scale." *Journal of Educational Psychology*, 1921, p. 399.

[3] "Mentally Defective Children," 1914, p. 13.

These inequalities of development . . . always produce a want of equilibrium, and this want is the differentiating attribute of the defective child." Doll[1] has raised this want of equilibrium to the importance of a pathognomonic sign of "potential feeble-mindedness." The component mental processes which determine the intellectual capacity develop uniformly in the former (the normal child) and not so uniformly in mental defectives. In the Binet tests the typical normal child has a basal year not more than one year below his chronological age and passes but a few tests beyond his chronological age. The potential feeble-minded, on the other hand, has the basal year more than one year below the chronological age or at least seriously below the total mental age rating and may have more than one basal year, that is to say, he 'scatters,' failing in tests one would expect a normal child of that age to pass, and succeeding in others not expected.

We have made a five-fold analysis of our data on scattering in relation to grade of intelligence, based on 2206 Binet records, and find little, if anything, in support of the above conclusions.

1. The greatest amount of scattering (based on the average amount of advance credit earned) is found among the normals in the Stanford and 1911 scales, and among the deferred in the 1908 scale, while the grade showing the least scattering is the imbeciles in the Stanford scale, the morons in the 1911 scale, and the normal in the 1908 scale.

2. If we confine the comparison to the extreme cases of scattering, we find that the proportion of subjects who earn 11 or more credits above the basal age (Table II), is greatest for the normal in the 1911 and Stanford scales, and for the deferred in the 1908, and least for the borderline in the 1911 scale, and for the imbeciles in the Stanford and 1908 scales.

3. Compared with the average amount of scattering found for all the subjects tested by each scale, it is found that in the Stanford scale the imbeciles and potential feeble-minded scatter less than the average (by 0.24 and 0.13 year, respectively), the morons slightly more (by 0.11 year), and the normals appreciably more (by 0.49) year. In the 1911 scale those who scatter less than the average are the imbeciles (by 0.13 year), the morons (by 0.13 year), and the borderline (by 0.03 year), while the normals scatter more (by 0.17 year). In the 1908 scale those who scatter less than the average are the imbeciles

[1] Preliminary Note on the Diagnosis of Potential Feeble-mindedness. *Training School Bulletin*, May, 1916.

TABLE II.—PERCENTAGE OF SUBJECTS PASSING ELEVEN OR MORE TESTS ABOVE THE BASE

	1908 scale			1911 scale			Stanford-Binet		
	Boys	Girls	Both	Boys	Girls	Both	Boys	Girls	Both
Normal	10.8	0.0	8.5	24.3	10.0	21.2	56.2	50.0	55.5
Retarded	7.5	14.2	8.9	15.1	14.2	14.9	39.1	0.0	31.0
Backward	3.6	14.8	6.2	18.2	8.1	15.6	43.4	34.7	42.3
Deferred	14.2	15.7	14.5	14.3	5.2	11.1	40.0	16.6	33.3
Borderline	3.7	12.5	5.7	6.1	4.1	5.7	44.3	41.1	43.6
Potential feeble-minded	....	....	....	....	....	....	36.6	15.3	30.9
Borderline and potential feeble-minded	....	....	....	....	....	....	42.5	34.0	40.4
Morons	9.6	21.5	13.8	6.4	9.8	7.6	43.0	44.0	43.4
Potential morons	....	....	....	....	....	....	46.9	44.9	46.0
Morons and potential morons[1]	....	....	....	....	....	....	44.7	44.4	44.6
Imbeciles	0.0	2.2	0.9	10.1	13.6	11.6	30.1	20.7	26.2
All	5.6	13.1	7.7	13.7	9.3	12.5	41.6	34.7	39.7

[1] In the 1908 and 1911 classifications the potential morons are included among the morons, and the potential feeble-minded among the borderline.

(by 0.02 year), the borderline (by 0.06 year), and the normals (by 0.25 year), while the morons scatter more (by 0.14 year).[2]

4. If we group together as *normal* all of those diagnosed as above borderline intelligence (*i.e.*, the normal, retarded and backward), as *subnormal* all those diagnosed as borderline and lower, and as *feeble-minded* the morons (including potential morons) and imbeciles, we find that in the 1911 scale the scattering is the least for the feeble-minded, and in the 1908 scale for the normals, while it is greatest for the normal group in the 1911 scale. The difference between the feeble-minded and normal group amounts to 0.16 year in both scales. The averages are the same for the subnormal and feeble-minded group in the 1908 scale, and for all groups in the Stanford (the difference amounting to only 0.01 year).

5. We may, finally, consider the number of ages from and including the base in which one or more tests were passed. The morons and potential morons pass tests in the greatest number of ages in the Stanford scale, the backward in the 1911, and the deferred in the 1908, while the retarded pass tests in the smallest number of ages in the

[2] The average amount of scattering for all the subjects is 1.02 years in the 1908 scale, 1.39 years in the 1911 and 1.63 years in the Stanford.

Stanford, (the "retarded" grading slightly under normal with an average IQ of 91), the imbeciles in the 1911, and the normals in the 1908.

It is evident from the foregoing that the conclusions would differ somewhat according to the scale followed. We believe, however, that some of the significant discrepancies may be cleared away by a study of, first, the average intelligence age level of the subjects in each diagnostic category in each scale and, second, by an analysis of the extensiveness of the testing of the subjects in each category. It is evident that groups of subjects with a high intelligence level could not scatter very much in the 1908 scale because of the small number of tests in ages 10 to 13, and the complete absence of tests above age 13. Moreover, in a previous comparison of scattering (in the 1908 scale) among subjects extensively tested and those sketchily tested, we found that the average number of advanced credits earned was invariably larger for the extensively tested group in every Binet-Simon age admitting of comparison. The pupils given a wide-range test earned from 0.75 of a year to 1.75 years more extra credits than those given a narrow-range test. Curiously the ratio of the number of tests passed to the number given was greater for the extensively tested than for the sketchily tested showing that a disproportionate amount of the credits came from the higher ages. On the basis of the experimental findings we were led to conclude that "the amount of credit

Table III.—Average Intelligence Age (Binet-Simon) of Subjects in Each Classification

	Stanford	1908 scale	1911 scale
Normal	9.34	9.26	8.77
Retarded	9.31	8.95	8.47
Backward	8.27	9.25	8.85
Deferred	5.93	7.07	6.62
Borderline	7.04	8.96	8.49
Potential feeble-minded	6.66		
Borderline and potential feeble-minded	6.95		
Morons	7.27	8.32	7.85
Potential morons	6.00		
Morons and potential morons	6.71		
Imbeciles	4.51	5.69	4.74
Average	6.92	8.45	7.95

earned depends upon the extent of the testing and not upon the grade of intelligence of the pupils."[1]

An analysis of the figures in Table III suggests the scattering among the normal pupils in the 1908 scale was small because of their high intelligence level (the same explanation would apply to the backward in this scale whose intelligence level was the same and who scattered less than the average). Forty-two per cent of the normals had a base of from X to XIII.

On the other hand, the extensiveness of the testing (see Table IV), helps to explain why the moron average of scattering is somewhat above the general average in the 1908 scale, and why scattering is greatest among the deferred in this scale, who also pass tests in the largest number of ages from and including the base. It helps to explain why the moron average is slightly above the general average in the Stanford scale, why the morons and potential morons in this scale pass tests in the largest number of ages from and above the base, and why the retarded (and backward) pass tests in the smallest number of ages.

Table IV.—Average Number of Tests Given above the Basal Age

	In 1908 scale	In Stanford-Binet
Normal	8.7	19.5
Retarded	11.6	14.0
Backward	12.1	19.8
Deferred	18.4	22.9
Borderline	13.4	22.7
Potential feeble-minded	....	23.8
Borderline and potential feeble-minded	....	23.0
Morons	15.6	24.8
Potential morons	....	28.2
Morons and potential morons	....	26.3
Imbeciles	18.6	23.8
All	13.8	22.8

From the above analyses we only seem justified in concluding that normal pupils scatter most and imbeciles least. Certainly there is no warrant for the assumption that "unequal development," "lack of uniformity," or "scattering" in intelligence is the "specific charac-

[1] *Journal of Delinquency*, 1917, 315f.

teristic," the "differentiating attribute," the pathognomonic sign of feeble-mindedness or potential feeble-mindedness. In fact, it seems to be the very want of irregularity, at least so far as general intelligence is concerned as determined by scattering in the Binet scale, which characterizes the lower grades of the feeble-minded. The conclusion here reached is in harmony with our earlier findings and with the findings of Pressey, who has studied irregularity among various types of subjects in the Binet scale, by different methods,[1] and Mathews, who has studied the problem in a group of delinquents.[2]

The Relation of Scattering to Intelligence Age Level

The smallest amount of scattering occurs in ages 2, 3, and 4, and in ages 10 to 12 in the 1908 scale. The data from the latter ages are probably unreliable because of the limits of the scale at the upper end. In the 1911 scale the corresponding ages are 2, 12, and 13 (the scattering being, again, artificially limited in the latter two ages because of the limited number of tests), and in the Stanford scale ages 3 and 10.[3] The greatest amount of scattering comes in ages 6 and 7 in the 1908 scale, in ages 11, 6, 9 and 10 in the 1911 scale, and in ages 11, 8 and 5 and 6 in the Stanford (ignoring ages above 11 because of the fewness of the subjects). The results are discrepant in the different scales, indicating, again, the danger of attempting to draw positive conclusions from one scale of tests. If we are justified in drawing any conclusions at all, it would be that the smallest amount of scattering occurs in the lowest mental age levels, and possibly that the largest amount tends to occur in the middle range of ages. This conclusion does not entirely harmonize with other findings. Pressey's tabulation shows that the scattering is about the same in the IQ range from

---

[1] Pressey, S. L.: Irregularity on a Mental Examination as a Measure of Its Reliability. *Psychological Clinic*, 1919, 236f.

Irregularity on a Psychological Examination as a Measure of Mental Deterioration. *Journal of Abnormal Psychology*, December, 1918.

The Distinctive Features in Psychological Test Measurements. *Journal of Abnormal Psychology*, 1917, 3f.

A Comparison of a Girls' Reform School, Attendants at a State Hospital for the Insane and Public School Children, by Means of Certain Tests of Intelligence. *Journal of Criminal Law and Criminology*, 1921, 258f.

[2] Mathews, Julia: Irregularity in Intelligence Tests of Delinquents. *Journal of Delinquency*, 1921, 355f.

[3] It should be pointed out that the differences referred to in this and other sections are sometimes almost negligibly small.

—76 to 125, and perceptibly higher in the range above 125.[1] On the other hand, in Mathew's data for delinquents the scattering is least for IQ's below 76 among the boys, and for IQ's below 76 and from 110 to 125 for girls.[2] Whatever significance, if any, differences of scattering may have in relation to intelligence age level, it is well to point out that the difference is quite marked between the age levels showing the least and the greatest amount of scattering, amounting to a year in the 1908 scale, a year and a third in the Stanford, and almost a year and two-thirds in the 1911. These differences are considerably greater than the corresponding differences between the intelligence categories.

The Relation of Scattering to Neurotic, Psychopathic and Delinquent Types

We shall compare (1) the average number of advance credits earned by each of these groups and by all of them combined into one group[3] with the corresponding average for all the subjects tested by each scale, for the combined group of normals ("normal," "retarded" and "backward") and for the limited group of normals, *i.e.*, those strictly diagnosed as normal; and (2) the percentage of subjects in each of these unstable groups and among all those tested and among those strictly diagnosed as normal who passed more than 10 advance tests in each scale.

Neurotics

1. The average amount of scattering for the neurotics was 0.01 year less than for all the subjects in the 1911 scale, and 0.24 year less than for all the subjects in the 1908 scale, but 0.17 year greater than for all the subjects in the Stanford scale. Compared with the combined normal group, the scattering was 0.16 year less for the neurotics in the 1908 scale, 0.06 year less in the 1911, and 0.19 year more in the Stanford. Compared with the limited normal group, the neurotics scattered 0.02 year more in the 1908 scale, 0.18 year less in the 1911 scale, and 0.31 year less in the Stanford.

2. Based on the percentage of subjects who passed 10 or more

[1] *Psychological Clinic,* 1919, p. 236.

[2] As before.

[3] Number of neurotics, 155; of psychopaths, 22; of delinquents, 352; total, 529. The diagnosis of these types was based on the physicians' reports, the psychological finding, and the personal history (social and school records). The delinquencies include such offenses as truancy, disorderliness, lying, stealing, and bad sex practices.

advance tests the ratio was 7.7 per cent less among the neurotics in the 1908 scale, but 9 per cent greater in the 1911 scale and 8.6 per cent greater in the Stanford. Compared with the limited normal group, the proportion for the neurotics was 8 per cent less in the 1908 scale, the same in the 1911 scale and 7.2 per cent less in the Stanford.

Psychopaths

1. The median amount of scattering among the psychopaths exceeds the general average by 0.10 year in the 1908 scale, by 0.03 year in the 1911 scale, and by 0.27 year in the Stanford. Compared with the combined normal group, the psychopathic excess amounts to 0.19 year in the 1908 scale, 0.04 year in the 1911, and 0.29 year in the Stanford. Compared with the limited normal group, there is a psychopathic excess in the 1908 scale amounting to 0.37 year, and a psychopathic deficiency in the 1911 and Stanford scales amounting to 0.14 year and 0.21 year, respectively.

2. Compared with all the subjects, the percentage who passed over 10 advance tests is greater for the psychopaths by 6.5 per cent in the 1908 scale, by 30.6 per cent in the 1911, and by 13.6 per cent in the Stanford scale. Compared with the limited normal group, the proportion among the psychopaths is 6.2 per cent and 21.8 per cent higher in the 1908 and 1911 scales, respectively, and 1.6 per cent less in the Stanford.

Delinquents

1. The average amount of scattering among the delinquents is 0.17 year less than among all the subjects in the 1908 scale, but exceeds the general averages by 0.04 year in the 1911 scale, and 0.03 year in the Stanford scale. Compared with the combined normal group, the delinquents scatter 0.09 year and 0.01 year less in the 1908 and 1911 scales, and 0.05 year more in the Stanford. Compared with the limited normal group, the delinquent average is 0.09 year greater in the 1908 scale, 0.13 year less in the 1911, and 0.45 year less in the Stanford.

2. Compared with the whole group, the proportion of delinquents who passed over 10 advance tests was 3.8 per cent less in the 1908 scale, but 5.1 per cent greater in the 1911 scale and 15.2 per cent greater in the Stanford. Compared with the limited normal group, the proportion among the delinquents was uniformly less in all the scales, amounting to 4.1 per cent in the 1908 scale, 3.7 per cent and 0.6 per cent in the Stanford.

Unstables

Many writers find it impossible to draw any clearly definable, unambiguous distinction between neurotics and psychopaths, while others are inclined to consider many, if not most, delinquents as psychopaths or neurotics. It will be of interest here to average the results for all these subjects and treat them as a single group of "unstables." When so treated we find: (1) the scattering is 0.17 year less among the unstables than among all the subjects in the 1908 scale, but 0.02 year more in the 1911 scale and 0.09 year more in the Stanford scale. Compared with the combined normal group, the scattering among the unstables is 0.09 year less in the 1908 scale, 0.02 year less in the 1911 scale, and 0.11 year more in the Stanford. Compared with the limited normal group, the unstable average is 0.09 year greater in the 1908 scale, but 0.14 year less in the 1911, and 0.39 year less in the Stanford. (2) The proportion of unstables who pass over 10 advance tests is 4.1 per cent less in the 1908 scale, but 6.5 per cent more in the 1911, and 12.5 per cent more in the Stanford, when compared with the figures for all the subjects. Compared with the limited normal group, the proportion is less for the unstables in all the scales, amounting to 4.4 per cent in the 1908 scale, 2.3 per cent in the 1911 and 3.3 per cent in the Stanford.

Here, again, the results differ according to the particular scale and according to the particular criterion employed. The average for the unstables exceeds that for the feeble-minded group; it possibly tends to exceed the general average slightly, although the tendency is by no means uniform; while among the unstable group the scattering is slightly greater for the psychopathic in nearly all the comparisons. But it is a question whether this peculiarity can be elevated to the rank of a pathognomonic sign of psychopathy, as has been done by Mateer,[1] who has accepted "more than four years of scattering above the basal year as an indication of psychopathy." We have found some individuals among all types—normal, subnormal, feeble-minded, delinquent, neurotic, psychopathic—who scatter very greatly, and others who scatter very little. Mathews records the same observation, "While many individuals who show great scattering in their tests are recognized as unstable, neurotic or psychopathic in their make-up, there are others in whom these tendencies are quite as pronounced who

[1] Mateer, Florence: The Future of Clinical Psychology. *Journal of Delinquency*, 1921, p. 283f.

give very even tests." Some recent writers have emphasized the importance of accompanying the mental age rating with a statement of the amount of scattering. That may be advisable, but the significance of little or much scattering has not yet been made clear by the studies which have thus far appeared. It is not yet certain whether scattering can be used as a pathognomonic sign of any type of mental defect, although our previous analyses seem to have shown that epileptic and psychotics as groups scatter more than any other groups—a conclusion apparently also reached more recently by Pressey. Nevertheless, we have examined epileptics and psychotics who scatter very little. Irregularity in intelligence development is not an indispensable condition or accompaniment of the epilepsies or the psychoses, although it is a frequent complication, possibly chiefly in the demented stages.

The Relation of Scattering to Sex

In the 1908 scale, the girls scatter more than the boys in all the diagnostic classifications except two, the difference between the averages for all the boys and girls amounting to 0.21 year. If the comparison is confined to those who do over 10 advance tests, the proportion is greater for the girls in all the classifications except one, while the proportion for the whole group who thus scatter is 7.5 per cent greater among the girls than among the boys. On the other hand, the boys earn more advance credits in every diagnostic classification in the 1911 and Stanford scales (using the complete figures), the boys' average amount of scattering exceeding that of the girls by 0.10 year in both scales. If we confine the comparison to the subjects passing more than 10 advance tests, the proportion is larger for the boys in all except two classifications in the 1911 scale, and in all except one in the Stanford. On the average, the proportion of boys who thus scatter exceeds the corresponding proportion of girls by 4.4 per cent in the 1911 scale and by 6.9 per cent in the Stanford. Thus the girls consistently vary more in the 1908 scale, and the boys in the other two, which are usually recognized as being more accurate than the 1908. This might be considered as slight confirmation of the view that the male sex is more variable than the female. Yeung,[1] in a recent examination of a small number of Chinese by the Stanford-Binet, found that the girls were only 0.63 as variable as the boys according to the Pearson coefficient of variability.

[1] Yeung, Kwok Tsuen: The Intelligence of Chinese Children in San Francisco and Vicinity. *Journal of Applied Psychology*, 1921, p. 267f.

Difference in Scattering in the Different Scales

Ordinarily we should suppose that the scale which shows the greatest amount of scattering is the most inaccurate. We should expect it to contain a larger number of improperly placed tests. It is, however, difficult to compare the scattering in the scales we have used because of the lack of tests in the upper ages of the 1911 scale, and especially the 1908 scale, because of the variation in the number of tests in the different ages in the 1908 scale, the difference in the method of obtaining the base in this scale, because of the larger number of tests, and the absence of tests in certain ages in the upper part of the Stanford scale, because of the unequal value of the tests in the Stanford scale above age 10, and because the extent of the testing may not have been equal in the different scales. Bearing these difficulties in mind, we find that the scattering is considerably greater in the Stanford scale, the average excess amounting to 0.24 year as compared with the 1911 scale, and 0.61 year as compared with the 1908. We have been aware of this peculiarity or weakness of the Stanford scale from our earliest experience with it. We have been struck by the unusually low bases which subjects frequently obtain in this scale because of failures on single tests in various ages which seem to be of more than average difficulty. It is frequently necessary to test all the way down to ages four or five in the case of subjects grading eight years or higher in intelligence, in consequence of which it is necessary to give a large number of tests in this scale in order to be fair to the subject. As a matter of fact, the testing was considerably more extensive in the Stanford than in the 1908 scale, for which we have comparable data, as indicated by the following figures, which show the number of tests given above the base in each category in each scale.

Table V.—Number of Tests Given above the Base in each Diagnostic Category

	Normal	Retarded	Backward	Deferred	Borderline	Morons	Imbeciles
1908	8.7	11.6	12.1	18.4	13.4	15.6	18.6
Stanford	19.5	14.0	19.8	22.9	23.6	24.8	23.8

A STUDY OF HIGH SCHOOL SPELLING MATERIAL II

JOHN A. LESTER

The Hill School, Pottstown, Pa.

(*Continued from February Journal*)

4. *Groupings for Purposes of Teaching.*—The reader will observe that the purpose of the investigation was not in the main to seek causes but to classify phenomena for the purpose of economical teaching. The aim was not to explain *why* these words are misspelled, but to observe *how* they are misspelled, and to group together the misspellings which exhibit a kindred nature. In certain cases, as for instance the fourth class in Table IV below, this classification indicates the cause at the root of the misspellings indicated. But, in general, the aim is to show the nature, degree, and percentage of certain forms of misspelling. The results of this classification are given in the form of a table. Each column for convenience is numbered and corresponding explanatory notes are added.

1. *Word-compounding.*—Under mistakes of word-compounding are classified all forms whether written solid, separate, or hyphenated (*twentyfive*, *him self*, *with-out*), as have no support in good usage, and such as plainly alter the meaning intended by the writer (*e.g.*, *a white bearded barber*). The prevalence of this kind of inaccuracy in writing is revealed only from accumulated records of errors made in free composition. But few of the mistakes in compound adjectives preceding their nouns (*e.g.*, *high powered engine*), though occurring with great frequency as a class, will occupy a high place in the arrangement of individual frequencies. They are individual, and occur in response to a special need. To determine the relative amount of this kind of misspelling, the sum total of the recorded mistakes in word-compounding was computed, compared with the sum total of recorded misspellings of all kinds, and found to constitute 15.9 per cent of the errors in spelling in free composition. This figure, 15.9, appearing in the table in parentheses, indicates the weight of this error more exactly than the percentage 6.2, which indicates the mistakes of this nature occurring in the most commonly misspelled 775 words.

A detailed examination of the mistakes of word-compounding shows that the tendency to separate is 15 times stronger than the tendency to combine in the writing of the students whose work is considered in this investigation.

TABLE IV.—CLASSIFICATION OF THE 10,853 TYPE MISSPELLINGS OF THE 775 WORDS MOST FREQUENTLY MISSPELLED

	1	2	3	4	5	6	7	8	9	10
Nature of the error	Word-compounding	Prefixes and suffixes	Confusion of words similar in sound or appearance	Mispronunciation	Apostrophe in possessives	Analogy	Doubling or singling of consonants	Final single silent *e* before a suffix	Mistakes which would be avoided by knowledge of some common Latin roots	Substitution of a vowel or vowel combination for one of similar sound
Percentages of the total misspellings	(15.9) 6.2	15.4	13.3	12.1	(8.2) 2.9	7.6	6.4	3.7	3.6	3.6

TABLE IV.—*Continued*

	11	12	13	14	15	16	17	18	19	20	21
Nature of the error	Final single consonant before a suffix	Substitution of a consonant or a consonant combination for one of similar sound	Internal modification: attraction, absorption, transposition	ei, ie	Capitalization	Omission or addition of a final silent *e*	Approximation to phonetic spelling	Apostrophe used for contraction	Final *y* before a suffix	Spellings in French form	Doubtful
Percentages of the total misspellings	3.4	3.4	3.4	3.2	(2.1) 0.4	2.0	1.4	1.3	1.3	0.9	4.5

2. *Mistakes in Prefixes and Suffixes.*—In this classification, misspellings in the *form* of prefixes and suffixes, denoting ignorance of what they are, were differentiated from misspellings made in the junction of prefixes and suffixes with the stem, indicating often ignorance of the general principles of word formation. Mistakes in the two noun suffixes *-ance* (*-ence*), *-er* (*-ar*, *-or*), and in the suffix *-ent* (*-ant*) account for more than half of all errors of this kind; suffixes are in general misspelled more often than prefixes; and the aggregate of mistakes in the form of suffixes and in the composition of prefixes amounts to 84 per cent of all the mistakes of this nature.

3. *Confusion of Similar Words.*—Under this head were classified mistakes not only in homonyms (*hear*, *here*), but in words which show such similarity of form as to lead to occasional substitution (*formally*, *formerly*). The great amount of misspelling attributable to the confusion of similar words is due to the high frequency of common words like *too*, *principal*, *there*, *led*, whose forms have not been firmly fixed in the minds of children in the elementary grades.

4. *Mispronunciation.*—There is a danger of classifying under this head any word whose misspelling appears to indicate a pronunciation other than the normal, but it is quite practicable to test the evidence of the records in the matter of pronunciation. Words whose common misspellings seem to indicate a common mispronunciation may be introduced into a typewritten piece of connected prose, and the pronunciation of the words observed when, without preparation, the prose is read aloud. This test proves for instance that while in 29 recorded misspellings of *thought*, 27 take the form of *though*, and all the recorded misspellings of *officer* take the form *officier*, neither *officier* nor *though* indicates a mispronunciation. On the other hand, it proves that the misspelling *atheletic* which occurs 32 times out of the recorded 35 misspellings of this word, is a true record of a common misspelling. This test was applied to all the words whose misspellings were in whole or part assigned to mispronunciation, and no word was included without evidence of (1), examples, few or many as the case was, of written forms indicating a mispronunciation; and (2), actual observation of such mispronunciation among students of seventeen to eighteen. The percentage indicates the amount of misspelling which is directly traceable to incorrect and slovenly habits of speech.

5. *Apostrophe in Possessives.*—Under this head is considered the omission, intrusion, or misplacement of the apostrophe denoting possession. It is clear, however, that no accurate measure of the

degree of this kind of misspelling in free composition can be obtained by taking its frequency in the 775 words most commonly misspelled. Hence all recorded misspellings of this class were computed, and the relative weight of this error (8.2) determined by comparing this total with the total of all misspellings of whatever kind.

But should such mistakes be regarded as misspellings rather than evidence of ignorance of grammar? Careful examination of the work of 560 candidates who wrote papers in 1918 and 1919 showed that 76 per cent of the books which contained errors in possessives also contained evidence elsewhere that the rules for the apostrophe to denote possession were known to the writers; and that the errors were therefore properly to be considered lapses, or errors of inattention.

Approximately one misspelling out of every twelve is a mistake in the form of the possessive.

6. *Analogy.*—This error differs from (3) above, in that here the form written is not a recognized word; the characteristic of the erroneous part of it bears an analogy to a recognized word or to part of one. In many cases this analogy produces reciprocal errors. For example, 85 per cent of the mistakes in *absence, sense,* take the forms *absense, sence;* 90 per cent of the mistakes of *always, all right* take the form *allways, alright;* 92 per cent of the mistakes in *choose, lose* take the forms *chose, loose.*

7. *Writing Single Consonants Double, Double Consonants Single.*—Under this head are classified mistakes of the nature indicated which do not fall into any other class.

8. *Final "e" Before a Suffix.*—Under this head are classified the mistakes in eliding or retaining a single silent final *e* before a suffix. The rule is commonly stated in three parts, as follows: A word ending in single silent *e* drops the *e* before a vowel. The silent *e* is retained before a consonant. The silent *e* is also retained to keep the soft sound of *c* and *g* before *a* and *o*. The percentage 3.7 indicates the applicability of this threefold rule.

9. *Common Latin Roots.*—Under this head are classified such misspellings as would obviously disappear in the light of a little accurate knowledge of Latin and of Latin roots. Arbitrary judgment was used in determining whether certain misspellings, such, for instance, as *comparitively*, should be placed under this head or another. These cases were always decided from the point of view of the practical teaching problem involved. The decreasing importance of Latin in the school curriculum was considered, and all cases such as the one men-

tioned (*comparitively*) which could be referred to both mispronunciation and to ignorance of Latin were referred to the former class. Under the head of ignorance of Latin might be placed many of the nouns and adjectives ending in *-ance, -ence, -ant, -ent.* But the same consideration of the meager knowledge of Latin which can be assumed to be practically available to the high school student in his spelling of English words, and the peculiar difficulties which characterize these endings, made it seem advisable to consider them separately. In other cases arbitrary judgment had to be used to decide whether a given misspelling (*e.g. innocent*) should be referred to mistakes in prefixes or to ignorance of Latin. Again where no knowledge of Latin can be assumed it is simpler to correct such an error by comparison with words similar in prefix, than with words similar in stem.

But when all such eliminations are made, there remain a number of misspellings which are most easily clarified by reference to the origin of the words in 42 common Latin roots. Examples are such forms as *benefit, ameteur, decend, descide, immagine, opperation, predjudice, volenteer.*

10. *Writing Another Vowel or Vowel Combination for One of Similar Sound.*—Under this head are classified mistakes of the nature indicated which do not fall into any other class. A number of misspellings of this general class are grouped with mispronunciations; others in which the prefixes (particularly *de-*, *re-*) or suffixes (particularly *-able*, *-ance*, *-al*, *-er*, *ious*) are written in incorrect form are grouped with mistakes of prefixes and suffixes; others are grouped with final *y*. Examples of the present group are such misspellings as *committy, privelege, endevor seveer, competetive, mineature, apperance.*

11. *Final Single Consonant before a Suffix.*—Under this head are classified mistakes of non-doubling or improper doubling of the consonant before the suffix which begins with a vowel. The common rule applicable in such cases may be expressed as follows. A monosyllable or a word accented on the last syllable, if it ends in one consonant preceded by one vowel, doubles the final consonant when a suffix beginning with a vowel is added. In other cases the consonant is not doubled. The rule applies with one exception in the list of 775 words.

12. *Substitution of a Consonant or a Consonant Combination for One of Similar Sound.*—From this classification is excluded as classified elsewhere mistakes of doubling or singling consonants; consonant mistakes in the junction of prefixes and suffixes with stems; several

mistakes of analogy; and several mistakes which would be removed by a knowledge of Latin roots.

Of the mistakes which were classified here, the writing of *sc* for *s* or *c* and the writing of *xs* for *x* (or vice versa) formed the largest group.

13. *Internal Modification: Attraction, Absorption, Transposition.*—Here were classed misspellings which show an expansion, contraction, or rearrangement of elements already existing in the proper form of the word. These internal changes may be classified as errors of attraction, in which an element already present either redundantly repeats itself, or else displaces another similar element (*accostom*, *athlectic*, *convienient*); errors of absorption, in which when two similar elements are present, one is assimilated to the other (*convient*, *irrestible*, *enthuiastic*); and errors of transposition, in which the elements suffer no expansion or contraction, but are merely rearranged (*entierly*, *Euorpean*, *gaurd*, *villian*). The amount of the misspellings due to each of these three processes is about the same.

14. "*ei*," "*ie.*"—Here are classified all misspellings which would be rectified by the mechanical application of the rule which follows. When the diagraph *ei*, *ie* is sounded like *a* in *fate*, or when, having the sound of *e* in *be*, it follows *c*, write *ei*. In other cases write *ie*. There are three exceptions to this rule in the list of 775 words.

15. *Capitalization.*—Mistakes in capitalization are naturally widely scattered. To obtain the relative weight of this error, the sum of all recorded mistakes in capitalization was compared with the total number of misspellings in the entire mass of writing. The most usual form of the mistake is the omission of the capital in the writing of proper adjectives (*french*, *indian*).

16. *Omission of Addition of a Final Silent "e."*—Examples of omission are such forms as *glimps*, *Main*, *medicin*, *practis*. Examples of addition are such forms as *ascende*, *dependente*, *prefere*, *stomache*. Mistakes of omission are the more numerous in the proportion of 3 to 2.

17. *Approximations to Phonetic Spelling.*—Here are classified attempts, as it were, at phonetic spelling, usually characterized by the omission of a silent letter, or the substitution of one which more nearly represents the sound. Examples are *cubboard*, *curtesy*, *evrybody*, *gard*, *Wensday*.

18. *Apostrophe in Contractions.*—The mistake of misplacing the apostrophe in contractions is more frequent than the mistake of omitting it.

19. *Final "y."*—Under this head are classified mistakes in final *y*

when a suffix follows, or when in plurals and verbs it is changed to *-ies, -ied*. The rule as commonly expressed has two parts, as follows. Final *y* preceded by a consonant becomes *-ies* in noun plurals, and *-ies*, *-ied* in verbs. Final *y* preceded by a consonant becomes *i* before any suffix which does not begin with *i*. There are no exceptions to this rule in the list of 775 words.

20. *Misspellings in French Form.*—A small number of words are frequently misspelled by the substitution of the French form of the word or an approximation thereto. The evidence for this classification is (1) that the French form predominates in the misspellings of these words, and (2) that the misspellings are often observed to occur after the study of French has begun. Seventy-seven per cent of the misspellings of *affairs* take the form *affaires;* 91 per cent of the misspellings of *minute* take the form *minuit* or *minuite;* all the recorded misspellings of *cover*, *officer*, take the forms *couver*, *officier*.

Of the formal rules usually presented in the teaching of spelling only four appear in the above classification. They occur in columns 8, 11, 14, and 19. Conjointly, they apply to some 12 per cent of the observed misspellings. Other rules usually taught were examined to determine their value in the teaching of this material. These rules were taken from four spelling books, one of which prints 12 rules, another 14 and the third, 27. Since none of the rules examined had an application amounting to more than one-half of one per cent of the misspellings actually made, they are not included in the classification which is likely to prove of use in the teaching of this material.

A report of the results obtained from a system of presentation and drill based on the foregoing study lies beyond the scope of this paper. The use of such a system under classroom conditions shows beyond question that greatly increased efficiency in the process of learning to spell may be obtained with an accompanying decrease in the time and labor expended.

Conclusions

1. The prevalence of lapses shows that the teacher of spelling must apply himself (1) to the formation of habits of accuracy and attention in writing and in reviewing what has been written; and (2) to the imparting of a body of knowledge.

2. Since some 25 per cent of the misspellings of high school and preparatory school graduates are misspellings of derivatives, the teaching of material drawn up on a dictionary basis will not be adequate.

3. If the critical points of the words to be presented are definitely known, and if the material is presented with insistent emphasis upon these critical points, at which most misspellings originate, a great gain in time and efficiency may be effected.

4. The rules usually taught have a very small ratio of applicability to the material to be learned. The four commonest rules, relating respectively to final silent *e* before a suffix, final single consonant before a suffix beginning with a vowel, *ei* and *ie*, and finally *y* before a suffix, cover conjointly less than 12 per cent of the misspellings recorded.

5. The five largest classes of misspellings, together including nearly 65 per cent of the whole are, in order of importance, the following: (1) mistakes in word-compounding, (2) mistakes originating in prefixes and suffixes, (3) confusions of words similar in sound or in appearance, (4) mistakes traceable to mispronunciation, (5) mistakes in the use of the apostrophe to denote possession.

6. The most direct means of gaining economy and efficiency in the teaching of spelling would appear to be twofold. (1) To teach material which the students concerned *do not know;* in other words to select material not on the basis of word frequencies in adult vocabularies, but on the basis of word frequencies in the misspellings actually occurring in free written composition. (2) To present that material with insistent emphasis on those critical points in the words presented which, in the material studied above, cause nearly 77 per cent of the entire mass of misspellings.

Books and Articles Referred to

Ayres, Leonard P.: "A Measuring Scale for Ability in Spelling." New York City Division of Education, Russell Sage Foundation, 1915.

Bawden, H. Heath: A Study of Lapses. *Psychological Review*, Monograph Supplements, Vol. III, No. 4.

Cook, W. A. and O'Shea, M. V.: "The Child and His Spelling." The Bobbs-Merrill Company, Indianapolis, 1914. (In Childhood and Youth Series, ed. M. V. O'Shea.)

Gregory, B. C.: The Rationale of Spelling. *Elementary School Teacher*, 8, pp. 40–45.

Hollingsworth, Leta B. and Winford, C. Amelia.: The Psychology of Special Disability in Spelling. New York City, Teachers College, Columbia University, 1918. (Teachers College, *Columbia University Contributions to Education*, No. 88.)

A COMPARATIVE STUDY OF A BORDER LINE DEFECTIVE AND A NORMAL CHILD OF THE SAME MENTAL AGE

LAURA REMER

Stanford University

Introductory note by L. M. Terman.—During the last five years, several students of educational psychology at Stanford University have made minor studies of the educability of children testing at various degrees below normal. One of these studies, the results of which are not yet ready for publication, involved a year of tutoring, two to three hours daily, of an 18-year-old girl of about 70 IQ. The study was made by Mrs. Gertrude Bell, a professor in the San Diego State Teachers College. Mrs. Bell did the tutoring herself, using the most ingenious methods she could devise. The purposes of the study were: (1) to find out how much the subject's school achievement could be improved, as measured by elaborate educational tests taken before and after treatment; and (2) to find out whether such intensive educational treatment would affect the IQ. The net results of the experiment showed a large improvement in subject matter achievement, amounting on the average to nearly two grades advance for the year. On the other hand, the IQ showed no change except for the slight improvement which would be expected to result from several repetitions of the test. Even specific coaching on types of material analogous to (but not identical with) that composing the Stanford-Binet scale had only negligible effect on the IQ.

The present study, by Miss Laura Remer, is considerably less extensive than that of Mrs. Bell, but it is offered for publication because of the scarcity of carefully made and accurately reported observations on individual cases of low IQ. Although not all subnormal children of a given IQ show the same degree of inability to master school subject matter, the case described in this article gives a concrete and fairly accurate indication of what may be expected educationally, from the typical 8-year-old having 6-year intelligence. Incidently, it gives a rather vivid picture of the differences between 100 IQ and 70 or 75 IQ. We need similar comparisons, even more searching, of the differences between 100 IQ and 140 IQ or higher.

LEWIS M. TERMAN.

Grace M. was 8 years and 7 months old when she was brought to Stanford University for a mental examination because she was making no progress in learning to read. She was given the Stanford Revision of the Binet-Simon tests with results as follows: Mental age, 6 years, 6 months; IQ 75. This placed her in the border-line group between dullness and feeble-mindedness.

She had started to school when she was 6½ years old. She was now beginning her third year in school and although her teacher was giving her extra help she was not learning to read. She had never been placed in a special group for backward children.

Physically, she appears quite normal. She has never had a serious illness. In personal appearance she is attractive, bright-eyed and often vivacious.

The mother is a Hungarian and claims to have come from a family of very good standing. She is extremely temperamental and easily excited to the point of hysteria. The father, an epileptic, is an uneducated American. There are four children all of whom, except Grace, test between 90 and 96 IQ.

The members of the family took a very unsympathetic attitude toward Grace. They were unable to control her in any way. In the neighborhood she was considered a vicious child. At school she could not study and would not let anyone around her study in peace. Habits of lying, stealing, and cunning deceit were becoming rapidly fixed.

In deciding what action to take the following facts were considered. Most children begin to learn to read when they are mentally 6 years old or younger. Grace was mentally 6 years and 6 months. Accordingly, it was arranged for the child to remain in school with the exception of one hour a day, during which she was to be tutored by the writer.

In order to compare her progress with that of a normal child of her own age, a child was selected from her class who by the Stanford Revision of the Binet-Simon tests was found to have an IQ of 100 and whose mental and chronological ages were 6 years, 6 months. This child, Ray S., received only the class instruction. The study extended over a period of 4 months but only 3 months of actual work was accomplished.

Before beginning to tutor Grace, an effort was made to find out how much she had gained during her two years of schooling. She read from memory several pages from different primers. Stories which she had heard many times she repeated verbatim; of others, which she had

heard fewer times, she gave the thought correctly, but not as written. The primer pictures aided materially this memorization. When asked to give a certain word she began at the beginning of the sentence and read word by word until she reached the required one. When the first words of each sentence were covered she made good guesses but was sure of none of the words except "see." This one word she recognized wherever she saw it. In phonics she was hopelessly confused. She knew the sound of many of the letters but could not connect the right sound with the right letter. She could count by 1's and 5's to 50; read and write numbers through 10; recognize silver pieces of money; and discriminate fairly well size, length and weight. She copied writing quite accurately, but did not know what she was writing. She enjoyed the social subjects, language, story and dramatization, music and drawing, although her standing in all was below the average.

It was seen at once that teaching this child to read must be a process of teaching her individual words. Her past work showed that she was able to memorize the story sentence by sentence, and that a large amount of repetition and drill had not been sufficient to teach her one word from another. After she had gained a small reading vocabulary the following experiment was tried in order to see how rapidly she memorized. A new short story, containing twelve sentences and having no illustrations, was selected. As it was in the nature of a review exercise, the story was not particularly interesting, so the content did not give undue aid. The first time Grace read it she asked for 22 different words out of 67, the entire number. Every second or third day the selection was read once, and no attempt was made to teach the unknown words. During the fourth reading she asked for only 5 words, and at the sixth reading she knew it perfectly. I then covered parts of the selection and pointed out one at a time the original 22 unknown words. She recognized none of them.

She had been taught reading chiefly by means of visual and auditory associations. Proceeding on the basis that increasing the number or kind of associations would increase the possibility of recall, we planned to increase gradually the kind of associations made with each word and note the results. But in order to have a check upon possible future progress, during the first 3 weeks of the tutoring only the two forms of association, visual and auditory, were used in the following manner. A simple story was selected from the primer which Grace chose. After I made sure by means of pictures, discussion and dramatization that the meaning of the story was clear, we concentrated

on a small unit of it. This unit was read and re-read, the sentences read in order and out of order, and a few phrases and words found and matched.

In order to show the method and the time necessary to teach one word, a brief review of one lesson near the beginning of the study will be given here. The paragraph which follows is a small unit of a well illustrated story in which the mother cat is trying to find food for her kittens.

> The cat saw a bird.
> The kittens saw it too.
> The bird saw the cat.
> It saw the kittens too.
> The bird flew away.

By reading the sentence she could name any word in the paragraph. With the first word of each sentence covered, she knew only those which we had previously studied. In this lesson she was learning "bird." She found the word as many times as she could in the paragraph, first with the entire paragraph exposed, then with the first word of each sentence covered. The word was written for her each time she found it. She then picked out the word from a small number of cards and matched it as many times as she could. This word was then put with several other review words and we played word-games for five minutes. At the end of this time she might or she might not recognize the word.

Her attitude toward the work was good. She wanted to learn, but her ability to concentrate varied greatly. Some days she would start to work with feverish concentration which lasted from 5 to 60 minutes. When a change of occupation was suggested she would beg to read "just the next page." On these days I could put all the words previously studied into a new story, the word arrangement entirely different, and she could read it easily. Never did she want to be told review words. Perhaps the next day she would be extremely flighty. Her attention could be held only by means of strong incentives and frequent change of occupation. On these days she remembered only the words on which we had drilled the longest length of time. This variability of concentration characterized all of her work throughout the 4 months.

In the first 3 weeks she learned 7 (!) very easy words, but in addition to this she began to understand the reading process. Formerly,

reading meant repeating something from memory or from imagination while she kept her eyes on the book. Now, she could not only differentiate between a few words, but she knew that she knew them.

It is generally recognized that a moderate amount of phonetic study improves the ability to read. However, if it is started too early or over-emphasized during the reading period it is apt to become an end in itself and block comprehension of thought.

In the case of Grace the phonetic training had been a hindrance rather than an aid in word recognition. She knew a number of the letter sounds, but with the exception of "S" could not associate the correct letter sound with the letter form. When asked to sound a word, as she was expected to do in her group work, she gave the first sound she thought of, regardless of letter form. Obviously, the thing to do was to discontinue all phonetic study until the point had been reached where it would be an aid rather than a source of confusion. But, because this was not possible in her group work, it seemed advisable to begin at the beginning and make the process as meaningful as possible, keeping it entirely separate from our reading period.

Consequently, each day we had a very short phonetic period beginning with games in slow pronunciation which were designed to help her hear correctly and recognize spoken sounds. Paralleling this oral work, the known sight word "run" was separated into its three sounds on the blackboard. She had little difficulty in locating and sounding the letters in order, but at the end of the three weeks she could not give the sound of "u" or "n" without beginning at the first of the word and sounding the letters in order. For instance, when "u" was shown her either on a separate card or in another word, she always went through this roundabout process: "It's in run, Run--r-u-n; *u*. Toward the end of the three weeks the known word "cat" was analyzed in the same way with the same results.

At the beginning of the fourth week, the following change was made in the reading period. Each day after the small unit of the story had been read, Grace chose one word for study. Then in addition to the visual and auditory associations, as described before, kinesthetic associations were emphasized. The word was pronounced and written for her several times on the board, while she watched. The work was then erased and she attempted to write the word. If she had difficulty, it was written for her again until she was able to reproduce it, but as her visual memory for form was quite good, this was not a tedious process. She then wrote the word as many times as she could in 5

minutes, or until her interest lagged, pronouncing it as she wrote it. The word was then, as before, put in the review list and used in games.

Although her ability varied greatly from day to day, it was evident that the addition of kinesthetic associations was beneficial. In case she could not recall a review word her response invariably was, "Don't tell me, I can get it if I write it." Although she sometimes wrote the word several times, if she did not recall it during the first or second trials, additional writings did not do so. When she could not recall it by writing I showed it to her in a familiar sentence; if this failed the word was pronounced for her.

During the 2 weeks in which the 3 types of association were made for every word, she studied 9 different words, 6 of which she learned comparatively easily and remembered. She was able to read by this time some 10 pages of a primer, the words of which she actually knew.

In phonics no new work was taken, but the kinesthetic association was made with each of the 6 letter sounds previously mentioned. At the beginning of each lesson it was necessary for her to go through the round-about process already described, then she remembered the sound until something else intervened. In the games, speed was gradually emphasized; but by the end of the fifth week, although she could give practically instantaneous responses to the four sounds, *r*, *n*, *c*, *t*, it seemed certain that the key word was inwardly vocalized before she was able to respond. Much improvement was noted in her ability to hear and give sounds correctly.

At the end of the first five weeks of tutoring, Grace was required to be out of school because of infectious skin trouble. Vacation came on and it was 4 weeks before her lessons could be resumed. During this time she was given no school help. She was glad to get back to school. Her teacher reported that she put forth more effort to learn and was easier to control than she had ever been.

I gave her a thorough review of everything she had studied with me, and found she had not forgotten. She read everything she knew before vacation and knew the individual words. She remembered every word in our word list with the exception of the last word studied.

During the rest of the time Grace was tutored (5 weeks) she received but 3 hours a week. Tactile associations were now added. The reading and phonic study continued in much the same manner as described. Short stories were chosen, for she soon tired of one and

wanted another. New stories were made on the blackboard using as far as possible her known vocabulary. She continued to study one word a day. The time spent on each word was gradually reduced to 4 or 5 minutes. A slow but steady progress could be seen. The combined associations, visual, auditory, muscular, and tactile added enjoyment and variety to the lesson which helped to hold attention and effort, and also reduced appreciably the length of time and the amount of mental stress necessary to learn a word.

The following table shows the ability of both Grace and Ray in reading and phonics; (*a*) when the study was begun, and (*b*) when it ended. The amount gained by Grace was found by averaging the results of thorough reviews given on 3 consecutive days.

When it was determined that 4 or 5 minutes was the least possible amount of time in which Grace could learn a word with a fair assurance that it would be remembered at least a few hours, I tried out Ray's ability. At 3 different times, 10 days apart, a list of words which he miscalled or did not know were taken from his reading. Five minutes directed study was given to each list which contained respectively 11, 12 and 14 words. Three days after each list was studied, I tested him and found that he remembered or was able to work out for himself 10 of the first list, all of the second, and 12 of the last. A little less than ½ minute was necessary to teach him a word.

Comparative Table to Show the Ability of the Two Children in Reading and Phonics When the Study was Begun and at the Close, Three Months Later.

Ability When the Study Began

Reading		Phonics	
Grace	Ray	Grace	Ray
None.	Ray read with ease the hardest stories from three different first readers of average difficulty. He grasped the thought easily with one reading.	Knew one consonant, s.	Knew all consonants and short vowel sounds. Could sound new one- and two-syllable words containing short vowel sounds.

Ability When the Study Ended

Reading		Phonics	
Grace	Ray	Grace	Ray
Has a reading vocabulary of 38 words. Reads the easiest parts of several primers. From 8 to 15 pages each.	Reads easily and with understanding in any of several second readers which he has not seen before.	Knows 13 sounds. Can sound one-syllable phonetic words which contain these sounds. Learns one sound or one word in a 5-minute directed study period.	Has learned the long vowel sounds and 21 phonograms. Can sound independently nearly any phonetic word found in a second reader of average difficulty. Learns 10 or 11 words in a 5-minute directed study period.

It will be seen by the comparative table that at the close of this study Grace had a reading vocabulary of 38 words. Three months after her tutoring stopped, I gave her a review of the work she had done with me and of the work she had done since that time. Of the 38 words she remembered 23. The forgotten words were mostly the unusual words which she had not had in her recent reading, such as *Christmas*, *dress*, etc. Her teacher was now giving her and another defective child a daily 10-minute reading period. They were using a very simple text and not attempting to keep up with the regular class. Under these conditions she was able to remember nearly ⅔ of the words because she was using them daily and was not forced too rapidly into difficult material.

Grace's language ability exceeds her reading ability. She is interested in what is going on about her, and particularly in out-of-door things. Her guinea pigs at home, the pups in the neighbor's yard, a bird she had seen that morning, and the endless changes brought about by the seasons are all objects of her observation and interest which she talks about freely to anyone who will listen. Yet she displays little thoughtful curiosity. Her questions are not of the usual type which want to know "why" and "how;" she thinks emotionally rather than intellectually.

In the reproduction of easy, familiar stories Grace does as well as Ray. The interpretations of both are good. With less familiar stories Grace does very poorly. She is very fond of hearing stories. I found it advantageous to start the hour's work by reading to her, for she never failed to calm down and listen quietly as soon as the story started. I always questioned her closely to see if she grasped the thought and found that she did. The books read to her were those classified for 7- and 8-year-old children. When she was one of a group of listeners, her attention and interest varied.

The Stanford Revision of the Binet-Simon Tests gave the following vocabulary scores for both children: In the first test Grace made a score of 10, and Ray a score of 19, neither passing the 8-year-old requirement, which is a score of 20. Five months later Grace scored 20 and Ray 26. The re-test showed no change in Grace's IQ.

The Thorndike Reading Scale A-2, X series, was given to both children in oral form as neither could read the lists of words. The tests were scored by the method used in the 1917 Nassau County Survey; *i.e.*, counting a score of one for every word correctly classified. Ray scored 45 and Grace 39.

Graded in "usual quality" on the Thorndike Scale for Handwriting of children, writing samples of both Grace and Ray ranged between 9.3 and 9.9. The norm for grade II is 7; grade V, 9.3; grade VI, 9.9. In speed Grace averaged 19 letters a minute and Ray 21. The second grade norm is 35 letters a minute.

Grace enjoys the drawing and handwork period more than any other in the school program. She occasionally does a good piece of work; but unless her teacher stands by her and insists that she follow directions she cannot hold to one idea long enough to finish it.

While this study was in progress Grace's conduct showed some improvement. Realizing that her bad habits were partly due to her home conditions, Grace's teacher and I worked with the mother and older sister, trying to show them how we handled her at school. Grace daily reported to me her conduct at home, and her reports were checked up. After the first few weeks her improvement was quite marked. She knew she would be called upon to account for her acts at a definite time and that untruthful statements would be discovered. At the close of the study she was no longer the outstanding character in the room. When objects were missing, she was of course suspected, but in the last 2 months of the study she was found guilty only once, and in that instance she returned the article before it was discovered

in her possession. As far as we knew she had not destroyed other people's property. She had not thrown herself into a passion—a former frequent habit—since the Christmas vacation.

However, we do not expect improvement in this field to be of a permanent nature. Her conduct will always vary with the kind and amount of supervision exercised over her.

In conclusion, the following points were found to be most helpful in teaching this defective child.

1. Special instruction was necessary. She could learn nothing in a group of normal children except auditory memory work. As the group instruction became more advanced, her mental confusion increased.

2. The necessity of planning each lesson step by step and presenting in each lesson only a very small amount of new material was apparent.

3. The daily review had to be more detailed and thorough than with normal children. Words not used regularly were forgotten.

4. Perhaps the most significant part of the study relating to the method of instruction was the result of using the kinesthetic associations in teaching each word. The visual and auditory associations by themselves were not effective, but when they all were used, with the emphasis upon the kinesthetic, the word was taught more quickly and with more assurance that it would be remembered.

It is obvious that Grace should not be in a class with normal children. It is not fair to her, to the rest of the class, or to the teacher. She is now 9 years old. In the course of the next 2 years, unless a change is made, she will become very restless and dissatisfied when the looked-for promotions do not come.

To protect society from an increase of her kind, the only place for this child is in an institution for the feeble-minded, where adequate supervision is assured. She has no special ability to be trained, but she could be taught to do the type of thing she enjoys most; *viz.*, out-of-door work.

JUDGING HANDWRITING

A Critical Weakness of the Thorndike Scale Revealed by a Comparative Study[1]

WILLARD W. BEATTY

Presidio Open Air School, San Francisco, California

Since the Thorndike, Ayres and Starch Scales for Measuring Quality of Handwriting have come into general use, several comparisons of their relative efficiency have been made. Rudolf Pintner, in The Journal of Educational Psychology for November, 1914, published a study which he thought proved a slight superiority for the Thorndike Scale. The month following, Truman Kelley demonstrated that through an error in computation the conclusions were erroneous, and that the superiority, though slight, really lay with the Ayres Scale. Leroy W. Sackett in *School and Society*, October 26, 1916, published a comparative study also tending to favor slightly the Ayres Scale, and F. S. Breed in *The Elementary School Journal*, February, 1918, showed by a similar study that the Gettysburg Edition of the Ayres Scale yielded more accurate results than the earlier three-slant form.

The upshot of the whole discussion has been that all three of the above mentioned scales have been demonstrated to yield almost the same results in practice—to the extent that fairly accurate constants have been derived for translating values on one scale into comparative values on the other.

The data here presented were secured during a similar study, which yielded results indicating that little if any superiority could be claimed for any particular scale judging from comparison of deviations in judgment. The graphing of the results of the scoring does indicate a definite weakness in the Thorndike Scale, which may be of interest.

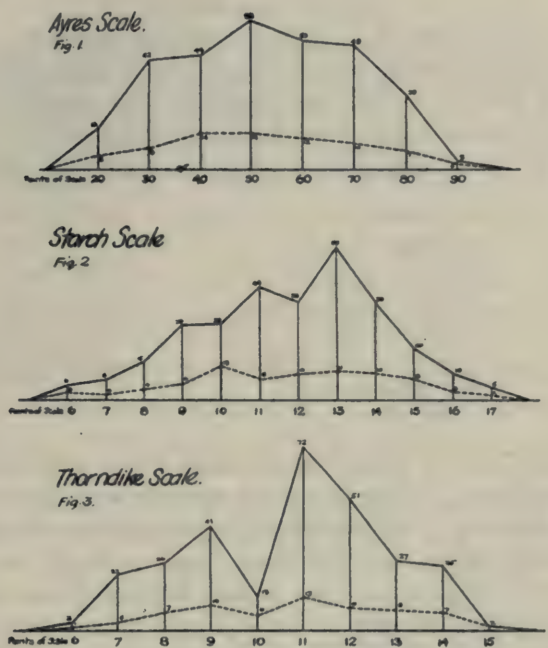
In the study, 17 typical handwriting samples were chosen from among 500. They were scored at intervals by 17 different persons, on each of three scales, the Ayres, Starch and Thorndike. After the remaining 483 papers were scored, 9 of the judges re-scored these 17 papers without knowing that they had been repeated. The results were tabulated, raised to a uniform scale for comparison and graphic charts prepared. As most of these data are very similar to those published by others, they are omitted. The method of graphing brings

[1]The data used in this article are drawn from a handwriting survey of the Daniel Webster School, San Francisco, carried out by a Seminar in Educational Measurements at the University of California, under Dr. Cyrus D. Mead, 1920.

out what would appear to be a fatal weakness of the Thorndike Scale, and is here presented.

Plate I shows the graphs. In all 289 judgments were made on each scale—17 papers, scored by each of 17 judges. These judgments were distributed over a certain number of points——that is, (Fig. 1)

Plate I



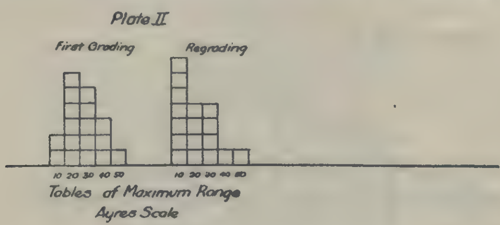
Curves Showing Frequency of Use of Points on the Ayres, Starch and Thorndike Scales.
Solid line - Number of Judgments at each point (total: 289)
Dotted line - No. papers with one or more judgments at each point.

out of a total of 289 judgments, point 20 on the Ayres Scale was used just 16 times, point 30 used 42 times, etc., as shown by the black line. These 16 scores of "20" were divided among 5 different papers; the 42 scores of "30" were shared by 8 different papers, and so on, as shown by the dotted line.

If the papers represented typical samples over a wide range of the scale these two curves should be fairly uniform for all three scales. Referring to the graphs, it is seen that Fig. 1, showing the distribution for the Ayres Scale is fairly uniform. Fig. 2 for the Starch Scale is a

little ragged, which may be accounted for by the fact that Starch offers half again as many samples in about the same maximum range. Fig. 3, for the Thorndike Scale reveals a remarkable drop at point 10, the judgments falling from a total of 41 at point 9, and 72 at point 11, to only 13 at point 10. The scores which might have been expected to fall at point 10 are divided between 9 and 11—11 getting the lion's share.

A glance at the Thorndike Scale reveals the difficulty. Only one sample of writing is used to illustrate point 10 as against two or three at every other point; and most unfortunately, this particular sample is a very peculiar and unusual type of almost back-hand writing. It is natural that in comparing samples with the scale, the judges should



seek an illustration of a type of handwriting which most nearly resembles the sample held. Point 10 cannot possibly be typical of 1 sample in 1000. In the long run, with many judges scoring each sample, and an average being struck, this deficiency would probably be overcome, but in any survey where one or two judges did most of the scoring, this unfortunate selection of point 10 added to the general unwieldiness of the scale, should eliminate the Thorndike Scale from consideration.

Plate II shows the result of practice in scoring with the Ayres Scale. The first diagram represents the maximum range on each of the 17 samples when first scored, the average of which will be seen to be about 27 points. The second diagram shows the change in maximum range on the re-grading, the average of which is a little over 21 points. It will be seen from the diagram that the number of papers where the range was over two points on the scale has risen from *two* on the original scoring to 7 on the re-grading—a "coming together" on the part of the judges which is quite remarkable.

Extended use of the scales, as well as the data presented, served to convince the judges that for convenience, reliability and all-round practicality, the Ayres Scale (Gettysburg Edition) is to be preferred for general use.

THE TORONTO MEETING OF THE SECTIONS ON PSYCHOLOGY AND EDUCATION OF THE AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

FRANK N. FREEMAN

University of Chicago

The recent meeting of the American Association for the Advancement of Science was the second one to be held in Toronto. In addition to a good attendance of representatives of psychology and education from the United States, there was a considerable attendance of Canadian psychologists and educators. The association between the workers from the two countries was very profitable, at least from the point of view of the visitors, and opened their eyes to the amount of scientific work which is being carried on in Canada. Some of this work was represented in papers which were read at the meeting by the Canadian scientists.

In the first session of Section I, two general papers were presented by men in the University of Toronto. J. A. Dale discussed the place of psychology in university curricula, and in his discussion represented particularly the point of view of the practical social sciences and more generally the demands of the applied sciences. He stressed the desirability of treating psychology in such a way that it would meet the needs of the groups engaged in practical enterprises. G. S. Brett discussed in a discriminating manner the various schools or movements, such as behaviorism, which have been represented in psychology of the twentieth century. W. B. Pillsbury also presented a paper of general nature. He brought the various points of view in modern psychology into opposition and raised the question whether these opposed points of view might be reconciled. He inclined to the belief that they represent permanent and perhaps temperamental differences, but that they are irrelevant from the point of view of progress of the science. A. P. Weiss discussed the fact of decreasing variability with increasing age, and speculated upon the possibility that society may prolong somewhat this period of plasticity.

As in recent years, many of the papers were devoted to mental tests. The results of testing children upon entering school and the divergence in ability in children of the same chronological age were reported upon by L. W. Cole. Several papers dealt with the technique of tests or with general problems relating to their value or particular

significance. The standardization of tests was discussed by Peter Sandiford. He emphasized particularly the necessity of restandardizing tests for use in Canada which are imported from the United States. Methods by which the validity of intelligence tests could be determined, largely on the basis of correlation, were discussed and illustrated by Raymond Franzen. R. M. Yerkes argued the necessity of considering psychological examination to be a much broader thing than intelligence testing, and maintained that it is necessary to use a more refined technique for research upon mental capacities than for application of intelligence tests.

A critical discussion of the application of intelligence tests to college students was presented by J. W. Bridges, who maintained that the prevailing correlations were so low that tests are of much less use in college for classification or other administrative use than is popularly thought. On the other hand, papers by G. M. Whipple and Wilhelmina Koerth described the application of intelligence tests to students in the universities of Michigan and Iowa. Whipple applied the tests to students on probation and concluded that other factors than intellectual ability were frequently the chief causes of failure. He therefore regarded lack of correlation as not necessarily evidence of the unreliability of the test. Miss Koerth showed that there was a wide divergence in the character of work done by those who stood in the upper and lower tenths in the intelligence test. A paper on the educational significance of mental tests, by William D. Tait, maintained that their results are to be accepted as indicating that education should be much more selective than it now is since there are many individuals who are incapable of profiting by it.

A number of papers on miscellaneous subjects may be mentioned. A paper on the psychology of the equation by E. L. Thorndike drew the distinction between two types of equation, the one being a mathematical statement which is made for the purpose of arriving at a solution, and the other being a statement of a relationship between a variable and one or more other variables. The first represents the common use of the equation in the conventional algebra, and the second represents its use in formulae, such as those which define certain types of graphic curves. The use of these two types is very apt to lead to confusion, and the author of the paper suggested various means by which this confusion might be avoided.

A paper by B. T. Baldwin and Lorle I. Steckher presented graphs showing the results of repeated mental tests and of repeated physical

tests of average and superior children. The results suggested that the mental growth curves of superior children diverge from those of average children, and that mental curves resemble and run parallel to physical curves. The inadequacy of children's concepts regarding matters which are assumed in their instruction was described and illustrated in a paper by Garry C. Myers. It was reported, for example, that many children who could recite the fact that Franklin was minister to France thought that this meant that he was a clergyman. A valuable study of the inmates of the Illinois penitentiaries was contributed by Hermann N. Adler. This study agreed with a recent study in Ohio and in general with the results of the tests in the army in showing that the inmates of penal institutions do not differ in general intelligence from the average of the population. The Army Alpha Test was used in these examinations. It was reported, however, that these prisoners in large numbers exhibited anomalies of behavior or of mental attitude. These anomalies, however, are not usually so serious as to make normal adjustment to society impossible. The somewhat related subject of psychiatry in the public schools was presented by Eric K. Clarke with illustrations from his experience in Toronto. T. R. Garth reported the results of a study which showed that pure blood Indians have different color preference from Whites.

One session was devoted to the applications of psychology, particularly in industry. The functions of an industrial relations department, with certain comments on the psychological aspects of the activities of such a department was presented by George W. Allen. The difficulties and psychological problems of job analysis were discussed by E. K. Strong, Jr. He pointed out that job analysis is the determination of the habits that a man must have to perform the work of the job and of the native qualifications that are necessary to enable the man to acquire these habits in a reasonable length of time. In the higher positions, however, the job is something more than the sum of its parts. A descriptive account of problems which confront the handicapped in securing and retaining work was given by Norman L. Burnett. He emphasized the factors of general attitude and personality in judging the prospects of success of a handicapped man and in placing him in a position. Alfred E. Lavell presented the successful experiment which is being made in Ontario to bring about a satisfactory attitude on the part of prisoners by means of employment outside the prison walls.

The writer was not able to attend two of the sessions of Section Q and therefore cannot report upon them.

The vice-presidential address before the section on psychology delivered by Dr. Strong, dealt with Control of Propaganda as a Psychological Problem. A psychological analysis of propaganda was made and supported by numerous illustrations. The paper emphasized the fact that propagandists influence the public by arousing sentiments and connecting these sentiments with particular forms of expression through suggestion. Our present methods of legal control are not adapted to this form of propaganda. Publicity does not seem to be a sufficient safeguard and an adequate solution of the problem is not yet at hand.

Dr. Judd, Vice-president of the section on Education, dealt in his address with the problem of the use of the scientific method in constructing the curriculum. He showed that there is at present no agency whose business it is to collect curriculum materials or to subject them to adequate standardization and drew illustrations from a recent attempt to develop and organize material for courses in the social studies. He offered no definite solution of the problem but presented it as one worthy of serious study.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY CECILE COLLOTON

Department of Educational Psychology, The Lincoln School of Teachers College

Intelligence and Educational Tests

What Los Angeles is Doing with the Results of Testing. Harlan C. Hines. Journal of Educational Research, 1922, January, 45–57. The use of the "intelligence survey" in organizing Ungraded and Adjustment Rooms in the elementary grades. Discussion of plans for a new marking system and the use of mental ability scores in educational and vocational guidance in the secondary schools.

Intelligence Tests in the Primary Grades. M. Edith Whitcomb. Journal of Educational Research, 1922, January, 58–61. The use of the Stanford-Binet in the grading and promoting of 2,360 primary children in Council Bluffs, Iowa. Comparison (in per cents) of quality of work and IQ's.

Intelligence Tests in Massachusetts Normal Schools. E. A. Kirkpatrick. School and Society, 1922, Jan. 14, 55–60. Results of administering the Thurstone Test to all students in the normal schools of Massachusetts. Comparison with scores of high school and college students.

Miscellaneous

A Diagnostic and Remedial Activity in Supervision. Bertha M. Rogers and Teresa Baker. Journal of Educational Research, 1922, January, 21–26. The use of the Woody Tests to show the need of improved instruction. Details of the experiment. One illustrative lesson used in remedial work is given in full.

The Correlation of Visual Memory and Perception of Perspective with Drawing Ability. Elmer Jones. School and Society, 1922, Feb. 11, 174–176. Preliminary report on an investigation being carried on at Northwestern University to determine native powers peculiar to children possessing art ability. Description of two tests for the special traits of visual memory and perception of perspective.

Problems and Solutions in Classification of School Children. Jennie B. Boyer. Detroit Journal of Education, 1921, November, 30–33. A general discussion covering retardation and its causes, individual differences, and the advantages of individual instruction.

Some Problems Arising in the Administration of a Department of Measurements. Helen Davis. Journal of Educational Research, 1922, January, 1–20. Discussion of problems such as Cooperation with Teachers, Test Administration, Classification, Publicity, etc., likely to be met by a director of measurements. Suggestions for principles and methods of solving such problems.

Provisions for Individual Differences in High School Organization and Administration. W. H. Hughes. Journal of Educational Research, 1922, January, 62–71. Are any provisions made in your school for individual differences of pupils? Summary of answers received from 221 high schools with an average enrollment of approximately 1000 each. Forty items are included in the tabulation; *e.g.*, basis of classification, supervised study, social recognition for superior students, etc.

The Basis of Individual Differences. Alfred Hall Quest. Detroit Journal of Education, 1921, November, 8. First section of an outline used as the basis of a series of lectures in the Detroit Teachers College. The outline will be continued in later issues.

Mind Set and Learning. Wm. H. Kilpatrick. Journal of Educational Method, 1921, December, 144–150. Continued from the November number. A popular presentation of the laws of learning.

Material in Four Delayed Issues of the Journal of Educational Psychology

Two Important Points with Regard to Age-grade Tables. S. L. Pressey. Journal of Educational Psychology, 1920, September, 355–360. A suggestion for the use of median age per grade and percentiles instead of per cent of retardation and acceleration, or age norms.

Superior Children—Their School Progress. Anna Gillingham. Journal of Educational Psychology, 1920, September, 327–347. Studies of gifted children in the Ethical Culture School and their achievement in school subjects. Descriptions of individual cases. Particular needs of the gifted child.

Solution of Problems in Geometry. Ben Wood and J. Carleton Bell. Journal of Educational Psychology, 1920, September, 316–326. An experiment to determine some of the chief factors in geometric ability. Descriptions of tests used and details of results.

Tests for Mental Alertness. L. W. Sackett. Journal of Educational Psychology, 1920, November, 430–444. Description of a test consisting of a short story and a set of exercises based on the story. Directions for administering, scoring, and interpreting results. Detailed tables.

Intelligence Ratings by Group Scales and by the Stanford-Binet Revision of the Binet Tests. G. M. Ruch and Lexie Strachan. Journal of Educational Psychology, 1920, November, 421–429. Correlation between scores on Army Alpha and Mental Age as determined by the Stanford Revision of the Binet-Simon Scale = 0.728; between Chicago and Stanford-Binet $r = 0.622$. Other correlations given with separate parts of the tests.

Tests for the Measurement of Certain Phases of Linguistic Organization in Sentences. Harry A. Greene. Journal of Educational Psychology, 1920, December, 517–525. Description of a language test for sentence organization.

A Comparison of Results Obtained by the Terman-Binet Tests and the Healy Picture Completion Test. E. B. Skaggs. Journal of Educational Psychology, 1920, October, 418–420. Extremely different ratings resulting from the Terman Binet and Healy Picture Completion Tests. Correlation = 0.15 ± 0.08.

Further Data on the Bell Chemistry Test. H. L. Gerry. Journal of Educational Psychology, 1920, October, 398–401. Results of the Bell Chemistry Tests in Biddeford, Maine, High School and Worcester Academy. Original standards too low.

The Correlation between College Grades and the Alpha Intelligence Tests. J. W. Bridges. Journal of Educational Psychology, 1920, October, 361–367. Results of the comparison of college grades and standing on the army tests of 486 students at Ohio State University. Four tables of interesting correlations.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

Books on New Mental Tests

Mental and Educational Tests in England.—Two recent contributons,[1] in the field of psychological testing show how steadily the movement is advancing in England. Mr. Burt's treatise is the most considerable contribution to the literature of the subject that has yet appeared in that country, and the preface is over-modest in setting forth its aim as that of presenting "a provisional set of practical scales for measuring intellectual ability and educational attainments" in the elementary schools. It not only does this, but also presents a vast amount of useful critical and statistical material. The scope of Mr. Godfrey Thomson's work is much narrower. It is an account of the group test which he devised in order to discover gifted children in localities of small educational opportunity.

The first section of Mr. Burt's book consists of an English version of the Binet-Simon scale, which was begun with the collaboration of M. Simon, and in its final version is the result of 8 years of experiment with over 3,500 children. The directions for giving and scoring each test are extremely careful, and there is an excellent key by which a test-score can be converted into a mental age in a few seconds. In discussing the various influences affecting proficiency in the tests, —schooling, sex, social status, etc.— Mr. Burt emphasizes the importance of schooling, and though the reciprocal relation between intelligence and educational attainments receives attention, he finally uses a partial-correlation equation to prove a proposition which will possibly be challenged, namely, that in determining the child's performance in the Binet-Simon scale, "intelligence can bestow but little more than half the share of school." In estimating the value of the scale for practical purposes he considers it greatest for use with defective children, less with normal children, and least of all for detecting the specially

[1] Burt, Cyril: "Mental and Scholastic Tests." London, P. S. King & Son, 1921, pp. XV + 432.

Thomson, Godfrey H.: The Northumberland Mental Tests. *British Journal of Psychology*, December, 1921.

gifted child. "The unwarranted claims advanced on its behalf by votaries in foreign quarters have among academic psychologists in this country become a commonplace and a byword." Yet as a provisional, workable substitute for a scientifically exact scale, he urges its use. His reason for standardizing for England the original Binet-Simon scale, instead of making a more or less drastic revision, as has been done in America and Italy, is that a present, uniformity can be secured and confusion avoided, only by this means. He hopes for a future English revision "which will ruthlessly abandon the tests which are now known to be worthless—sex, surname, date, months, two lines, comparing faces,—" and will consistently pursue the principle that has been fitfully attempted in America, the "internal grading" of each test, so that it will appear at the different age-levels with regular increments of difficulty—the method which Terman has adopted in the case of his vocabulary test, and that of repeating digits backwards. As a useful rough method for testing new tests, he suggests the use of Yule's colligation coefficient ω, the Coefficient of Association; an Appendix describes this statistical device in detail.

Besides the Binet scale, Mr. Burt gives one form of each of the usual intelligence tests, standardized with London children, Synonyms, Definitions, Instructions, Completion, Absurdities, Maze, and as an example of the thoroughness of his methods, we may point to the 16 criteria he employed in compiling his list of opposites (p. 223). It is interesting to remember that Mr. Burt was the first psychologist to use what is commonly known as the "Analogies" test to measure intelligence, when he was working in Liverpool in 1911, though Woodworth had analysed the mental processes involved in the recognition of such relationships 3 years earlier. Mr. Burt gave the test its name from Aristotle's ἀναλογια, proportion.

The intelligence test which he considers the most efficient, especially for older and brighter children, is his individual test of Reasoning Ability, which consists of 17 brief questions, preceded by the data necessary for answering them by ordinary logical inference. Each problem is assigned to the age at which 50 per cent of the children answered it satisfactorily. This test was used successfully last year in a small experiment in a Public School in Brooklyn.

The section of the book also gives some interesting graphs of the Distribution of Intelligence among London children, both in the ordinary elementary schools and the "Special" schools for Mental Defectives. After much discussion of this highly debatable point,

Mr. Burt suggests a Mental Ratio (IQ) of 67–70 as a theoretical line of demarcation for mental deficiency in children, *i.e.* to justify removal to a Special M.D. School. As provisional borderlines at the upper end of the scale he suggests that a Mental Ratio above 115 (or even 120) denotes central school ability, (Central Schools are lower secondary schools, usually having a commercial or industrial bias,) and that one above 130 (or 135) merits a full secondary school course. If this last estimate really represents the lowest level of intelligence of the "free-placers" in secondary schools, it shows the futility of comparing the average intellectual attainment of an American High School with that of a London Secondary School. In no public elementary school did Mr. Burt find an IQ higher then 160, though in a private school he tested a boy of 7 with one of 190, as high as that of the recently reported American and Scotch precocities.

The last section of the book is devoted to Educational Tests and Scales for the elementary subjects, with a description of individual cases of disability in each. The author gives norms for each age, but adds a warning against the danger that lurks in thus giving prominence to the average or median score. Just as an official minimum wage tends to become the maximum, or at least to limit it, so a risk arises lest better performances should tend to be depressed towards the mean. He would be startled to hear the devotees of standardization acclaiming this as an advantage. The Drawing scale consists of a median specimen of a drawing of a man for each age from 3 to 14. One wonders whether it is only a curious coincidence that in 5 of the specimens the man is smoking!

The whole book gives evidence of the scholarship of the author, and leaves one with a strong impression of his intense interest in the personality of the individual child.

For nearly a quarter of a century, poverty, unless very dire, has not excluded really gifted children from full educational opportunity in England, provided that they lived in or near a town. The ideal has been the "ladder" whose rungs are scholarships, sometimes with maintenance grants, leading from the Public Elementary School through Secondary Schools of various types, and perhaps a Provincial University, to the best that Oxford and Cambridge can offer. The scholarships to secondary schools are given to children aged about 11, who are presented for examination in Arithmetic and composition by the elementary schools, and it has recently been noticed by one Education Authority, the County of Northumberland, on the Scottish

Border, that more than one quarter of its schools present no candidates at all, and that these schools are chiefly found in the remoter country districts. It was felt that two reforms were necessary to help the gifted children in these districts, (1) full maintenance must be offered in addition to free schooling, (2) the examination must be in the form of a mental test which would give an equal or almost an equal chance to an unprepared child. Mr. Thomson undertook the task of preparing such an examination, and after considerable preliminary experiment he chose the following group tests: New Tests—(1) Hindustani Test, (2) Extra Number Test—Partly New Tests, (3) Middle Word Test, (4) Schema Test, (both based on the work of Prof. Stern of Hamburg, but new in the form of presentation) Old Tests, (5) Extra Word Test, (6) Number Series Test.

In choosing a large proportion of new tests, he says he was actuated in part by a desire to make coaching for the tests almost impossible, and in part by a desire to make the experiment as original a contribution as possible, and to increase the supply of mental tests. Much reliance has been placed on some of the tests he definitely rejected, *e.g.* the Completion test, but the common use of this as a teaching device in English schools would bar its inclusion in this particular series.

The new tests are interesting, particularly the Hindustani test, which may prove of special diagnostic value. Each is given in full in the article.

The tests were given to nearly 3000 children in addition to the scholarship candidates whom the rural schools sent in, and age-norms were tentatively fixed from the results. The scores seem to correlate satisfactorily with the Binet testing of the same children, which is still in progress. The highest IQ found by the new tests (174), was that of an 8-year-old boy in a small village in the heart of the Cheviot Hills; and a comparison of the distribution of intelligence in schools in different localities showed that the most intelligent children are to be found in the small country schools in the hills, the results there being even better than those in a well-to-do town suburb. Experiments in other parts of England have shown that children in large cities are on the average a year in advance of those in the country. But in this investigation, the only very large town, Newcastle, was omitted, as it has its own Education Authority. Moreover the Cheviot children are to a large extent the descendants not so much of farm laborers as of Border troopers. Mr. Thomson draws the tentative conclusion

that the highest ability is to be found close to cities and far away from cities, the intermediate areas being drained by selection. It will be interesting to discover whether this holds good for other countries too.

EDITH NEWCOMB
Teachers College.

3. *A Further Revision and Extension of the Binet-Simon Scale.*[1]—This manual presents a much revised and extended Binet-Simon scale. Its aim is to make it "possible for a teacher, a student of psychology, a parent, a worker in a juvenile court or in an institution for defectives or delinquents, to use the Kuhlmann scale readily and accurately after the reading of this volume."

The revision is based on the results of seven years of continuous work with about 7000 children and adults in the Minnesota School for the Feeble Minded, and in the public schools of Minnesota.

The manual consists of five chapters and an appendix. Chapter 1 describes the nature of the revision and extension of the scale including a brief discussion of the general accuracy of the method and the significance of the results.

Chapter 2 is a detailed discussion of the general principles of the year scale based on the three topics (1) the requirements of the individual test; (2) the construction of a year scale; and (3) the establishment of norms.

Chapter 3 gives complete information as to the best way to conduct an examination, how to use the scale in abbreviated form, and how to determine the mental age.

Chapter 4 describes the materials needed for the examination and presents detailed directions for administering the tests and scoring the responses.

Chapter 5 summarizes various comments on the individual tests.

The appendix consists of a table of intelligence quotients which "gives all quotients from 25 to 150 for the ages of three years to maturity and the mental ages of 3 to 15 inclusive." Dr. Kuhlmann says his tests show the maximum mental age to be 15 years. "Maturity" evidently means 15 years in chronological age also for the table of chronological ages stops at 15 years 0 months.

Of the various revisions of the original Binet-Simon Scale probably

[1] Kuhlman, F.: "A Handbook of Mental Tests." Warwick & York, Inc., Baltimore, 1922, p. 208.

the best known is Terman's, which is called the Stanford Revision of the Binet-Simon Scale, or more popularly, the Stanford-Binet. The outstanding differences between the Terman and Kuhlmann revisions are the increase in the number of tests from six to eight in each age group above two years, the extension of the scale at the lower end (Kuhlmann's scale begins with tests for the age of three months) and the use of a different method of scoring when the subject passes several tests in the highest age group of the scale. The same tests are given in years 13, 14 and 15 but they are scored differently for the different ages.

The aim of the author has been to give a larger place to actual performance as compared with verbal response and to insure a higher degree of accuracy in scoring the results of an examination. Therefore, practically all of the new tests consist of things to do such as counting dots in a square; tapping blocks in irregular order; spelling familiar words backwards; crossing out q, r, s, t, in a pied test; following directions in a confusing text; drawing upright forms in inverted positions. They are scored in terms of the time it takes to do the test and the number of errors that are made.

Examiners will be interested in using this scale and comparing its results with those secured from previous revisions. It is generally conceded that most of the revisions do not adequately measure mental age at the upper end of the scale and if this new scale proves the claims of its author in this regard it will be of distinct service.

CECILE COLLOTON
The Lincoln School of Teachers College.

---

4. *A Manual for Case Study* which is extremely compact yet definite and complete has been recently published by the California Bureau of Juvenile Research.[1] Comprehensive instructions, with sample data, are given for taking case histories of intelligence, temperament, physical condition, moral character, conduct, amusements, education, home conditions, etc., etc. Chapters are devoted to the scope and meaning of social case investigation, methods (such as interviews, correspondence, inspection of records, etc.), methods of evaluating data, the use of charts, symbols, etc. It contains tables or norms of height, weight, complete sample histories and excellent lists

[1] Williams, J. Harold and others: Whittier Social Case History Manual, Whittier, Cal. California Bureau of Juvenile Research. *Bulletin* No. 10, December, 1921, p. 98.

of references on each topic discussed. This booklet, which can be purchased for a quarter, merits use in normal school and college classes and could be studied profitably by educational workers in general.

A. I. G.

---

5. *Psychology Coming into Its Own in the Making of Children's Books.*—If you were stranded with no books for children to read from and had to write your own, how would you go about it? Very likely you would let the children write the stories which would make up their reading books. You would put into the books the things to which children of the different ages give spontaneous attention. You would like to be with the children throughout their waking hours, constantly tabulating and interpreting what they did and said. You would answer their requests for stories by telling them some, trying first one kind and then another; and living close to the children, studying and recording what they like and demanded you would not go far wrong.

You would find, for example, that 3-year-olds simply revelled in the mere recital in story form of the, to them, exciting things of the day, but to us the humdrum routine of life. When they, themselves, are in the center of the stage they are consumingly interested in the daily tasks of getting up in the morning, dressing, eating, playing throughout the day and the like. And if you did write books for children this way and then studied critically the books which children now read in the primary schools, you would be impressed by the fact that someone had sold the primary pedagogues a good sized gold brick; for the said pedagogues have resorted thoroughly to Mother Goose, fairy stories, repetitional animal and similar stories.

Now, this very sort of thing is exactly what Mrs. Lucy Sprague Mitchell and her colleagues have learned by living with the children in the City and Country School of New York City. Mrs. Mitchell's *Here and Now Story Book*[1] is the result. Really, this book is quite revolutionary. No doubt, it will fearfully upset our writers of Primers and School Readers, 50-foot shelves of which are still being published each year. If they will only read Mrs. Mitchell's introduction, they will at least dimly see the why and wherefore of all the change, and then, if they will trail a little group of children for a few days—one would be quite enough—and concentrate very hard upon what the children really did and said, and use a good pinch of Woodworth's,

[1] E. P. Dutton & Co., New York, 1921, p. 360.

Thorndike's and Dewey's Psychology on it all—why they would, this reviewer humbly believes, go and do likewise. No teacher of primary children should fail to read the introduction to this book. And thousands of homes should turn to the stories in it to read to children: The Dinner Hour, The Grocery Man; The Journey; How the Engine Learned the Knowing Song; The Fog Boat Story; Hammer and Saw and Plane; The Skyscraper; Things that Loved the Lake, etc.

H. O. R.

A New Hand Book of Modern Education for Teachers.—In making a curriculum for school children one constantly needs to know the evidence which has been collected on mooted points of social needs, what children learn, what their interests are, what we know about their heredity, their intelligence and their growth. One of the essential tools of the progressive and experimentalist in school practice is an up-to-date, complete and described bibliography of what biology, psychology and sociology has to say about the nature of the child and the educative process. It would be still more helpful if liberal quotations were supplied of what our leaders of prestige have to say about these matters. Finally, the teacher needs definite suggestions concerning books and materials which she can use with children.

If I were a teacher with these needs I would get Miss Gertrude Hartman's new book *The Child and His School.*[1] It does those things and very well, indeed. *First,* it gives the teacher a good bibliography of books and materials (850 of them) that teachers can select from to be read by children in the lower and intermediate grades. These references are classified to fit: (1) Community study (food, shelter, clothing, transportation, communication, conservation of wealth, education, recreation, religion, government and primitive life); (II) National life (in general, its government, its history) ; (III)The study of other nations. True, many of the books listed are very poor readings for children, but, as we who are collecting and organizing reading materials know, they are all we have. This is an excellent list to have at hand.

Second, it does summarize, quote from and interpret for the teacher the scientific basis of education. This book is an example of what the Bureau of Educational Experiments (of New York City) is doing. It combines the use of a broad philosophic interest in child life and the

---

[1] E. P. Dutton & Co., New York, 1922, pp. XI + 248.

improvement of society with the scientific foundations of these in biology, psychology and sociology. It does not turn up its nose at the notion of measuring the child's ability to spell or that he must master certain arithmetical tools that will be needed in life. Most free educational institutions have displayed this attitude in recent years. Paralleling this characteristic of the Bureau to keep pace with the development of scientific education, they are committed to the working out (in schools) of the basic philosophy of the "free educationists" laid out with such clear vision by John Dewey. And Miss Hartman has made a very good illustration of this point of view in her book. She has also combined with it a keen use of scientific education.

H. O. R.

A New Statistical Book for Research Workers.—The scientific study of education waited on two things: *one*, the development of methods and devices of measurement; *two*, the taking over from biometry and mathematical statistics of the technique of statistical treatment of data. The publication of Professor Thorndike's *Mental and Social Measurements* in 1903 was epoch-marking. From that date on, for ten years, hundreds of technical publications in education appeared and rarely indeed, it was that footnote references to his book did not abound in the publications. My own *Statistical Methods Applied to Education* (1917) doubtless served to interpret on an elementary plane, methods by which the school man could treat his data. Neither of these two books, however, supplies the mathematical foundation nor the refined methods which research students in psychology and education need for the extension of their very scientific work. American students have been forced to used the elaborate publications of Karl Pearson, Yule, Elderton, Edgeworth and others. Professor Jones' new book,[1] especially part II, will help fill this gap admirably. Part I (using economic and biological illustrations) gives the mathematical basis of much of what is already available in American books on measurement, variables and frequency distribution; classification and tabulation of materials; averages, dispersion, graphs and correlation. American students will regret that Professor Jones has not discussed partial correlation and that they will still be forced to go to Yule's detailed original publication. In Part II students will have

[1] "A First Course in Statistics." G. Bell & Sons, London, England, 1921, pp. IX + 286.

access to a treatment of probability, sampling, curve fitting and type distribution curves that will be very helpful, indeed. There is little doubt but that we are entering upon the second stage in the use of statistics by students of education. The "statistics is arithmetic" stage in which we used frequency, averages, dispersion and correlation, blindly, is practically passed. The more high-brow stage of using statistics for the discovery of scientific law and of prediction we are now just entering upon. The work of Kelly, Otis, Ruml, Rosenow and others provides important applications in the determination of the reliability of tests and in the study of "law" in learning, growth and the like. Schools of education will need at hand, therefore, in increasing proportion such technical tools as Professor Jones' book provides.

H. O. R.

Another Manual on the Technique of Teaching.—Each of the new books which I have brought together for review for this month makes an addition to our method of teaching—some very large, some small. The next book to be discussed can hardly be said to do so. Davis' *Technique of Teaching*[1] is but another of a long stream of such books written largely out of the experience and observation of the author and only partially in touch with the newest scientific and philosophic developments. It contains a chapter of general principles of teaching, followed by a series of chapters, one devoted to the teaching of each of the principal school subjects—spelling, reading and literature, composition and grammar, arithmetic, history and geography. The best thing in the book is the set of exercises at the close of each chapter. Beyond this I find little to justify the publication of another general book in this field. In dealing with the separate subjects little or no reference is made to the psychological facts now available from the results of measurement. To one's astonishment he finds reference in the bibliographies at the end of the chapters to almost none of the scientific analyses of learning in the different "subjects." Instead the author refers prospective teachers to other books in the same category as his own—empirical and *a priori* pedagogies. The book reflects acquaintance with pedagogical theories of 1910, but not with the virile dynamic psychology and the scientific method of attacking educational problems which are developing so fast in 1922.

H. O. R.

[1] Macmillan, New York, 1922, pp. VIII + 346.

9. *A Novel Approach to Reading.*[1]—Recent contributions to the literature of method reflect the unsettled state of educational thinking. One hesitates to be duly critical of constructive contributions because of the good they may do. But if, by over- or under-emphasis, they tend to obscure real values and create new problems, criticism is urgent.

Miss Watkins presents a method by means of which beginners may be taught silent reading. She puts that method forth succinctly and systematically. She makes the materials available so that anyone may attempt to do what she has found possible by actual trial.

We quote from Chapter I:

"Using the Silent Reading Method, each pupil in a class of forty, long before the end of the first school year, carried out without hesitation the following printed commands in the presence of the class and visitors:

'Tell that man sitting by the window that the spinning wheel over in the corner is older than the telephone, electric light, railroads and the United States.'

'Draw an oblong, a circle and a square. Put your age under the circle, then go and turn on the electric light that is over my table.'

'The man standing by the door and wearing a gray suit lives in Omaha. Go to him, shake hands with him, tell him your name and your father's name; then show by the number of your swings on the swing just how old you are.'

'Ask the lady with the blue dress to whisper to you the time she came to Iowa City, then show us the time on the large dial. Then ask her if she came from Denver on the Rock Island Railroad. Tell us what she says.'"

Miss Watkins then asks and answers the following question:

"How can an immature and entirely untrained mind be taught in the short space of the first school year to grasp fully and execute commands involving a knowledge of every part of speech and of so varied an assortment of actions?

"The result is achieved by treating the child's intelligence as a full grown intelligence which simply requires to be informed, in a logical manner and without interposition of obstructive methods.

"There is apparently no limit to the capacity for learning which the child's mind possesses, save the limitation of time and circumstances. There is no reason why a child should not learn anything valuable which it is capable of learning in as short a time as possible."

Surely Miss Watkins would not have us select materials, methods and vocabulary on so arbitrary a basis! Should the first grade child's reading vocabulary include such terms as "personal history," "hibernate," "salutation," "doilies," "abdomen," "objects," and "pro-

[1] Watkins, Emma: "How to Teach Silent Reading to Beginners." School Project Series. J. B. Lippincott Company, Philadelphia and London, 1922, p. 133.

ject?" Some of these words are used so infrequently that they are not found among the 10,000 words listed in the Thorndike word counts. Is it feasible to introduce new word symbols in categories? Is an action response sufficient guarantee of the meaningfulness of a symbol or is it possible that the symbol acts as a direct cue to a particular response, regardless of real meaning? Is there a danger of word sophistication and reading facility without sufficient background in experience? What does the following sign mean to a 6-year-old? "Please do not ask for credit." Does the child who responds to the flash-card "vulgar," by saying its opposite, have any clear notion of what it is all about?

The directness of the method has much to recommend it. The use of flash-cards is no doubt conducive to concentration. The action response is valuable. Children *do* learn by noting differences.

The method is highly stimulating. The boy who was "so proud of his work and achievement that he took his tightly rolled test to bed with him" worries us. The satisfactions are not those which usually engross six-year-olds. The psychology of the method is fundamentally at fault in another respect. Miss Watkins speaks of the child's mind as though it were a receptacle. Another conception of learning makes for a more critical evaluation of motives and procedures.

Fortunately we read Dr. Ernest Horn's introduction in time to realize that the proposed method is not supposed to displace "instruction in literary appreciation" and extensive reading from printed materials.

L. Z.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Vol. XIII April, 1922 No. 4

THE RELATION BETWEEN MENTAL AND PHYSICAL GROWTH

BIRD T. BALDWIN

Iowa Child Welfare Research Station, Iowa City, Iowa

A few years ago, I published a study on Physical Growth and School Progress,[1] based on individual growth curves of children. The chief value of these curves consisted in the fact that they were the first to follow consecutively the same groups of children for several years in physical growth, school standing, and the relation of the two. Since the first report I have emphasized in various subsequent investigations the *individualizing method.* This paper presents additional empirical data on the analysis and significance of physical growth curves, the interpretation of similar data on mental growth curves, and the relation between mental and physical growth.

Physical Growth Curves

At present I have data at hand on approximately 2500 individuals for 30 physical traits with consecutive measurements on nude children for periods of from 18 to 24 semi-annual intervals. These boys and girls have had systematic medical inspection, directed play, physical training, and those falling outside a normal growth zone on account of disease history or abnormal growth have been eliminated. If one of these specific physical traits is selected, growth in standing height for example, the curves show a number of individual characteristics and definite basic principles of growth for different types of children.[2]

For growth in height, the curves show that the boys are as a rule taller than the girls, except from approximately 11½ to 13½ years of age, the girls reaching their final height earlier than the boys. For

[1] Published by U. S. Bureau of Education, 1914, No. 10, pp. 215.

[2] Baldwin, B. T.: The Physical Growth of Children from Birth to Maturity. *Iowa Child Welfare Studies*, No. 1, 1921 (1), pp. 412.

both boys and girls the curves fan out, showing a wider range of distribution as the chronological ages increase from 4 to 18 years. The pubescent acceleration, when present, appears earlier for the girls than for the boys, usually preceded by a slight retardation. At the adolescent acceleration the curves approximate in appearance a series of concentric arcs, with the acceleration appearing earlier in the taller children (the upper arch) than in the shorter. Since the curves assume in general a railroad appearance, each boy and girl holds his or her relative position in the group for the ages from 4 to 18, with little crossing of the curves. The maintaining of the same relative position of individuals within a group permits for the first time the scientific approach to the problem of physical growth in height from the standpoint of *prediction*.

Prediction in Physical Growth

The degree to which prediction is possible measures in a very definite way the development and practical application of a science. For example, the prediction of physical and mental growth enables one to determine whether children are advancing at a normal rate. If it is found that they are not, remedial measures may be taken to accelerate growth or prevent over-stimulation. The significance of any increment of growth, physical or mental, depends fundamentally on what the status at later periods should be, the size of the increment being conditioned by the physical or mental type of the individual. Tall, medium and short children grow differently, with characteristic physiological stages of maturation which later affect the rate and completion of growth but do not affect relative ranking within a group.

This large series of long-time individual growth curves for different types of children enables us to predict what will be the later status of children who have been measured at the early ages. One method of doing this is to identify a child with one of the types of curves and estimate what his later development would be. Also, since the increments for growth from one chronological age to the next or between definite stages of physiological ages have been computed for a wide range of type growth curves, we can predict what should be the increment at any later period for a child who has been measured only a few times. So again, knowing that a boy of the average type at 7 years of age has gained 70.3 per cent of his final growth at 17 years of age, and a girl 74.2 per cent of her final growth, we can predict what the ultimate stature of a similiar child should be. These three methods

GROWTH CURVES in HEIGHT

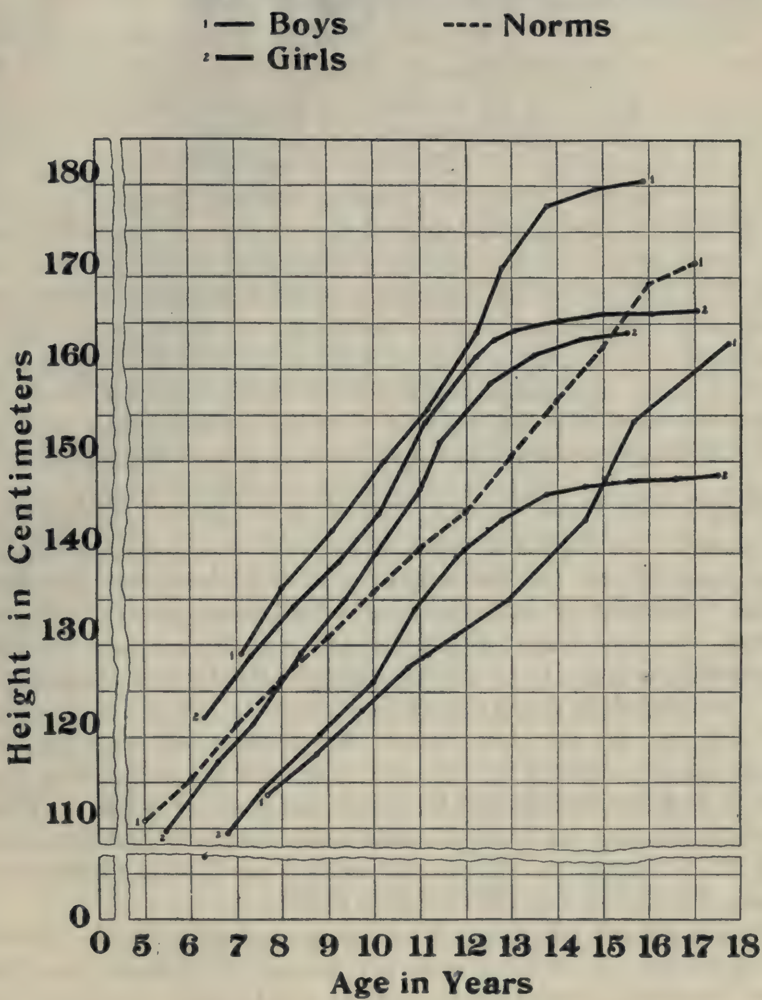


CHART I

of prediction are based on a knowledge of the child's chronological and physiological ages and the type to which it belongs.

The evenness of physical growth in height has been indicated by the parallelism of the curves as cited above. Further evidence of this evenness of growth is to be found in the subsequent rankings of individuals in the group by the *method of correlation.* This method gives us the rankings of individuals in the earlier and later ages within the group. Using the Pearson method, the coefficients of correlation for 125 boys and girls at 6 or 9 years of age and 6 years later range from +0.735 to +0.944. These high coefficients show that there is a great probability that a boy or girl who ranks tall at 6 years of age will also rank tall at 12 years of age, or, on the other hand, a boy or girl who is short at 9 or 10 years of age will be short at 15 or 16 years of age. Similar high correlations exist between birth and later ages as shown by a limited number of curves extending into adult life.

From the above correlation data one can also predict by using the regression formula (found in Yule) $y_1 - y = r\frac{\sigma y}{\sigma x}(x_1 - x)$ the height, for example, for individual cases at 12 years of age, from the height at 6 years of age, and the height for 15 or 16 from that at 6 years earlier. The PE of estimate on any individual case in these two groups, when the height of 12 year old boys is predicted from the height at 6, is found to be 2.98 cm., and for the 12 year old girls 2.58 cm. For the older group the PE of estimate is 2.09 cm. for the boys and 2.81 cm. for the girls; that is, the chances are even that any measurement predicted for six years later from the height at the age of 9 or 10 years for girls will lie within the limits of +2.81. In this case, the chances are 1 to 4.5 that a measurement will lie outside of 2 PE or ±5.62 cm., that is, the chances are 8198 in 10,000 that a predicted measurement will be within 2 PE.

A detailed analysis of this kind is now possible on physical growth curves because we have a sufficient number of children who have been measured consecutively under standardized conditions by uniform methods and by a uniform scale. No such complete data exist on mental development. Nevertheless, with certain material recently obtained it is possible to follow the general method that has been outlined. The detailed analysis of these data is given in a study[1] now in press by B. T. Baldwin and Lorle I. Stecher.

[1] The Mental Growth Curves of Average and Superior Children. *Iowa Studies in Child Welfare,* No. 1, Vol. II, 1922, pp. 59. (In press.)

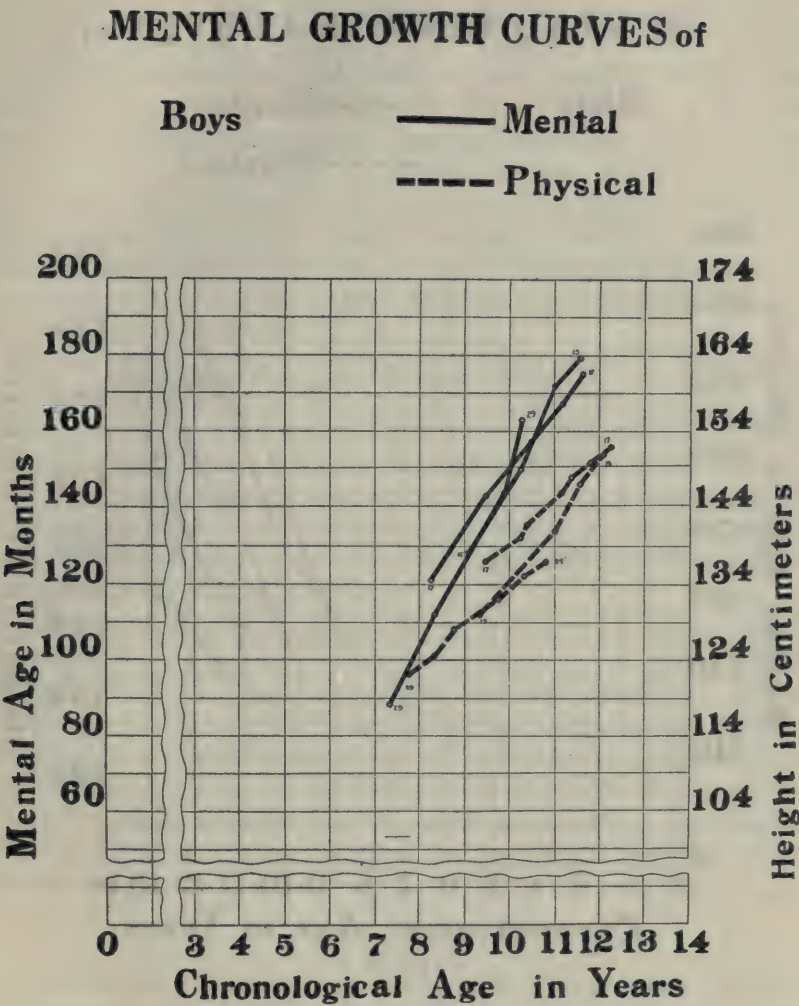

MENTAL GROWTH CURVES of
Boys
Mental
Physical
200
180
160
140
120
100
80
60
0
174
164
154
144
134
124
114
104
Mental Age in Months
Height in Centimeters
0 3 4 5 6 7 8 9 10 11 12 13 14
Chronological Age in Years

CHART II (*a*)

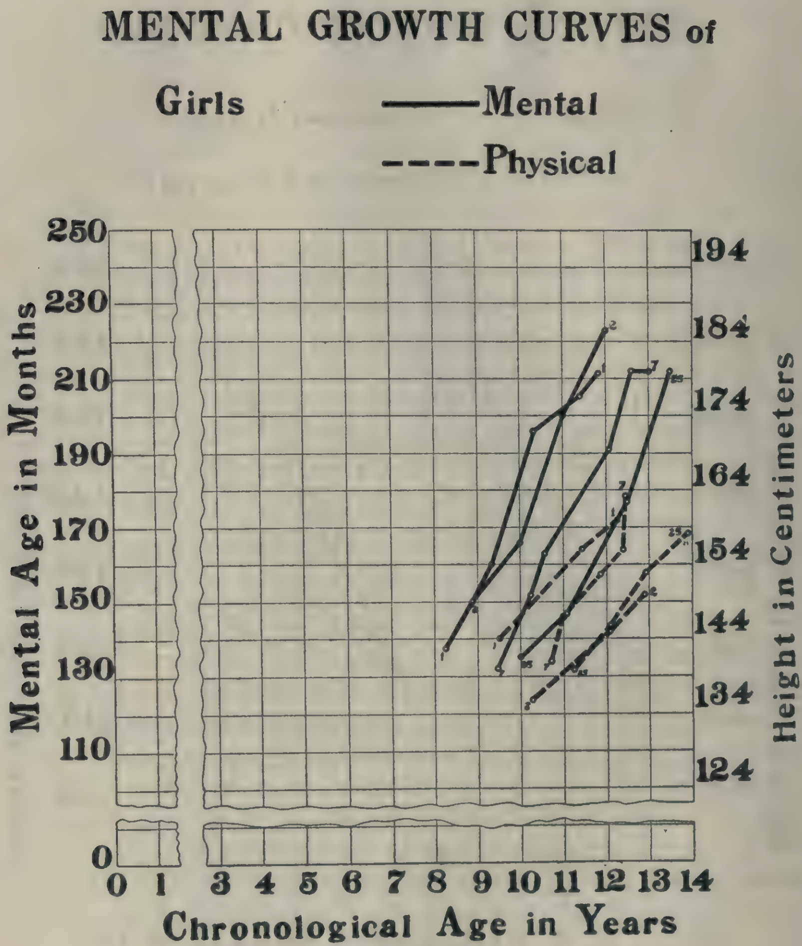

MENTAL GROWTH CURVES of
Girls
Mental
Physical
Mental Age in Months
Height in Centimeters
250
230
210
190
170
150
130
110
0
194
184
174
164
154
144
134
124
0 1 3 4 5 6 7 8 9 10 11 12 13 14
Chronological Age in Years

CHART II (*b*)

Mental Growth Curves

In 1917, several hundred children were examined at the Iowa Child Welfare Research Station by the Stanford Revision of the Binet Scale, with a view to obtaining subsequent individual mental growth curves. On account of war conditions and the shifting of school population, the number of cases for this purpose has been reduced to 143, of whom 42 have had 4 examinations and 36 additional cases have had 5 consecutive examinations. The individual mental growth curves have been plotted. Chart II gives a few individual physical and mental growth curves.

Chart III shows the mean mental growth curves of average and superior boys and girls. The mental examinations on which these growth curves are based were made at irregular intervals. In order to plot the curve at the customary one year intervals, the mental ages, in place of being assigned to the year nearest the chronological age, as is usually done, were re-calculated by using directly the rate of improvement which had existed just before and after the chronological age included within this interval.

The curve does not extend beyond 14 years. These superior and average children develop at different levels and grow increasingly dissimilar with age. This divergence in growth curves of average and superior children has been assumed as probable, but has not here-to-fore been empirically demonstrated. There is, it will be noted, a change in the trend in the curve at the approach of adolescence. This is especially noticeable in the curve for the girls. The superior children show this acceleration about one year earlier than the average children.

Prediction in Mental Growth

Consequently, if we know the growth history of a sufficient number of children, we can tell from these individual curves what should be the later status of children of the same type who have received only the earlier measurements. The results in Chart IV also show that the IQ is only approximately constant upon successive examination. The girls present a greater irregularity in development than the boys in this group. Using again the method of correlation for predicting relative ranking in mental growth, the coefficient between the first IQ and the fifth IQ is +0.82 for the boys and girls, which indicates that they keep their relative positons after an interval of approximately 4 years. For larger groups with fewer repeated exam-

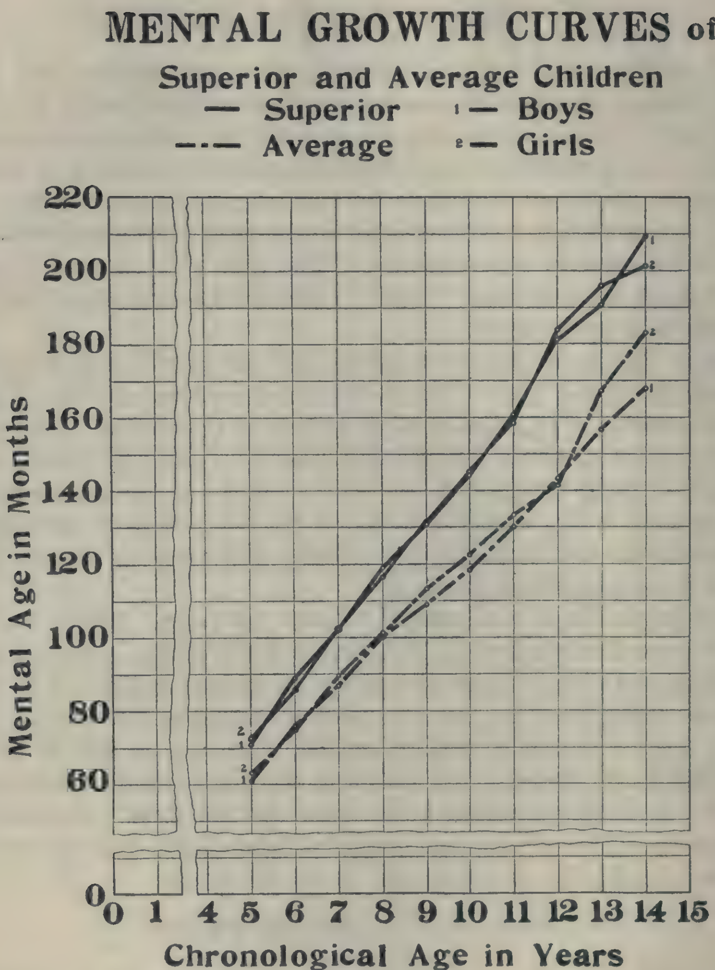

MENTAL GROWTH CURVES of
Superior and Average Children
— Superior
-·— Average
1 — Boys
2 — Girls
220
200
180
160
140
120
100
80
60
0
Mental Age in Months
0 1 4 5 6 7 8 9 10 11 12 13 14 15
Chronological Age in Years

Chart III

inations, the coefficients range from +0.69 to +0.93, the near-lying examinations having higher coefficients as a rule. In most instances two examiners did all of the testing; in a few, four, but the number of examiners has little effect on the correlation.

From the four coefficients of correlation involved for the five series of IQ's, the PE of estimate obtained by means of the regression equation for the prediction series of the second from the first IQ equals 4.2 points; the third from the first, 7 points; the fourth from the first, 6.2 points; and the fifth from the first, 5.5 points. The mean time intervals were 13, 28, 36 and 41 months respectively.

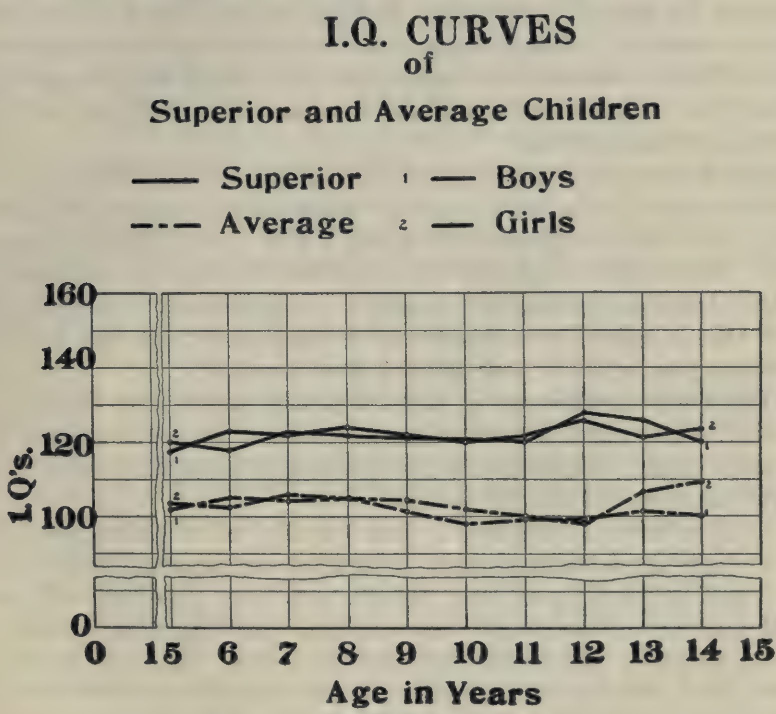


CHART IV

It is not possible from the data at hand to make an exact determination of the amount of error of prediction for various intervals of examination, since all of the children in this group have had repeated measurements, which influence, no doubt through practice, the size of the correlation for the longer intervals. In order to determine how accurately one may predict a child's IQ at one year or two years later,

the correlations will have to be obtained on a sufficient number of children at each examination interval without intervening periods. In the absence of such long time data, one can say that it is possible to predict a child's IQ with a PE of from 4 to 7 points.

The Relation Between Physical and Mental Growth

It is apparent from the charts that certain phenomena associated with physical development show themselves in a decrease or increase in the mental age and in the IQ at certain chronological ages. Both superior boys and girls show a rise in the mental age and in the IQ's between the ages of 11 and 12. Average girls also show this adolescent acceleration, although it appears later than in the case of superior girls. The IQ curve and the mental growth curve of the average boys do not show this phenomenon, possibly because they have not reached the stage of acceleration.

The general pre-pubertal increase in mental development becomes evident earlier in the case of superior children than in average children and in the case of superior girls about a year earlier than in average girls. These same contrasts exist between average and accelerated boys and girls physiologically classified. In general all of these curves show that in regard to these adolescent phenomena boys and girls are usually a year apart in their general development.

The mean mental age of physiologically accelerated children is higher than for physiologically retarded children when those above the norms in height and weight (the accelerated) are compared with those above the mean mental age for each year.

In the earlier monograph it was stated that if pedagogical age be accepted as a fair equivalent to mental development, tall, heavy boys and girls with good lung capacity are older physiologically and further along in their stages toward mental maturity, as evidenced by school progress, than short, light boys and girls. This conclusion is based on 21,682 final term grades and 5000 physical measurements on 125 boys and girls from the Horace Mann School at Teachers College, Columbia, University, and The Francis Parker School in Chicago.

Since the first study of the inter-relation between physical and mental growth, as shown by school progress, this work has been continued by taking the measurements of height, weight and total area of carpal bones of the same children, who were examined by the Stanford Revision of the Binet Scale at Iowa City. The correlations between these physical and mental traits for 49 girls, for example, are, for

height and mental age +0.89; for weight and mental age +0.71; for exposed area of carpal bones and mental age +0.83 as shown in the Baldwin-Stecher Study.

The size of the correlations is unduly influenced by the range of the ages of the children of this group, which is from 5 to 15 years. The correlations for the physical traits previously stated are based on relatively large numbers of children of the same chronological ages. The correlations between IQ's are probably not subject to criticism from this point of view, since the IQ compensates in a measure for the differences in the chronological ages.

Growth in height shows in all of our studies a high correlation with physiological age, or stages of physiological maturity, when various criteria of this age are used. On the average, weight correlates with height +0.809 for boys and +0.603 for girls at any two subsequent ages from 7 to 17. The inter-correlations of physical traits are consistently higher for boys; the coefficients of variability are also higher for boys.

The coefficient for the group of 49 girls by the method of partial correlation (with age constant) is between height and weight +0.57; between height and x-ray +0.52. For height and mental age the coefficient is +0.53. That is, physiologically accelerated girls (and the same holds true of boys) are mentally accelerated. The mental acceleration includes both stages of mental maturity and brightness, since the scale does not differentiate between these two distinct phases of mental growth. As I pointed out in 1914, no scale can accurately measure mental growth which does not take into consideration individual differences in physiological age for children of the same sex and the same chronological age. A thorough-going program for investigating the relation between mental and physical growth would require consecutive mental and physical examinations on a large group of children at regular intervals, with uniform methods and standardized scales for mental and physical maturity and intelligence. This would necessitate a tri-dimensional scale in place of our mental chronological age scales. The day for the study of physical growth processes by single measurements is over. They can tell us little or nothing that is not already known. The day for basing a study of *mental growth processes* on snap-shot cross-section group or individual examinations will soon be over. Let us begin to do what should have been done long ago, *i.e.*, plan consistently to make *intensive consecutive studies throughout a series of years on the same individuals.*

PITFALLS IN RATING SCHEMES

FREDERIC B. KNIGHT

Psychology Department, State University of Iowa

AND

RAYMOND H. FRANZEN

Research Director, Des Moines Public Schools

When subjects rate themselves or others for several qualities, certain tendencies (probably unconscious) are present which make it necessary to be on one's guard in the interpretation of the ratings. Two such tendencies will be reported in this article. Results of introspection generally, and ratings are of value only when these are interpreted as behavior and not as truthful accounts. If "A" reports that he thinks something, *that* is important irrespective of whether he really thinks it or not. The conflict between introspective and behavioristic psychology is not in the data obtained, but in the way the data are interpreted.

I. Pitfalls in Rating Schemes

In the ratings of an individual upon himself in comparison with others in a group, there is a marked tendency of the individual to overrate himself. This as far as we know is commonly assumed theoretically but actual data on the point are scant.

In the spring of 1921 we had 110 junior students in a university rate in *order of importance to themselves* a list of interests varying in character from the essential to the trivial. Samples of these interests are,—pleasing one's parents, dancing, the movies, development of character, magazine reading, the church, dress, and health. There were 34 interests in all.

The setting up of an order of interest on a basis of the pooled returns was perfectly feasible. Such an order of interests may be thought of as the relative importance of interests of typical juniors *when the ratings are taken introspectively*, or the order of interests of a typical junior *as he thinks he is*. We then had the same juniors rate these interests for the ideal junior. This then is the order of interests as the junior thinks it should be. The correlation between these two orders of interests was +0.46.[1] There is, then, a tendency for the junior to think he is as he should be. Later we had the same group rate these interests "in order of importance to the typical junior." The pooled

[1] Pearson Correlation Coefficient.

rating from these data may be thought of as the relative importance of interests of the typical junior when the ratings are taken objectively, or the order of interests in the junior as he is seen by his classmates. The correlation between this order of interests and the "ideal one" was -0.64. There is, then, a tendency for the junior to think his fellow-juniors are *not* what he believes they should be. The correlation between the order of interests taken introspectively and that taken objectively was $+0.13$. Then there is very little association between what the junior thinks he is and what others think he is. He is a very bad judge of himself. Thus students place themselves between the ideal and the typical fellow student. There is a higher association between what they believe they are and what they would like to be than between what they believe they are and what their fellow students believe they are. Students when rating themselves are, in their own opinion, nearer the ideal in their order of interests than they think their fellow students are. But, when 110 juniors rate their own interests, the resulting pool *is* that of the typical junior. Thus the very low correlation between the order of interest taken introspectively and the order taken objectively cannot be accounted for by difference of facts. It is due to difference in the point of view of the rater. Each believes he is to some extent the ideal of the group but he does not extend this courtesy to his fellows. There is in self-rating, for junior students at least, a marked tendency to overrate themselves or underrate their fellows, which amounts to the same thing as far as comparisons are concerned. Undoubtedly there is a self-defense mechanism at work in self-ratings. Just what the fluctuations of this tendency are is unknown.

We think that this tendency to overrate one's self and the extent to which any one individual does it has possible diagnostic value quite apart from the truth or falsity of the ratings themselves.

The following orders of interest have been obtained from each individual:

1. His order of interests as he envisages them.
2. The order of interests he believes to be typical of the group of which he is a member.
3. The order of interests which he believes to be ideal for himself.

The corresponding orders of interest obtained for the group by pooling those of the members of the group are:

4. The average order of introspective origin.
5. The typical order as the group judges it.

6. The ideal order the group judges it.

Many valuable diagnoses of traits are yielded by the intercorrelations of these orders and by the partials. The most important of these are:

(*A*) The amount and direction of the correlation between 1 and 2 indicate the attitude of the individual on the question of whether or not he likes what others like. If the order of interests as an individual believes them to be in himself correlates positively and highly with the order he believes to be typical he judges himself to be a normal human being. If this correlation is negative and high, then he believes himself to be "different." The value of knowing an individual who judges himself as being different or "funny that way" is of obvious import to the diagnosis of mental peculiarity and mental disease.

(*B*) The amount and direction of the correlation between 1 and 3 indicate the degree to which he believes himself to be what he wants himself to be. A high positive correlation between a person's ideal order and what he reports as his own actual order might mean a well-satisfied person, *i.e.*, he believes his values to be in the order of importance which constitute his ideal. If we could further determine that his interests were actually not in this order by finding how much he went to the movies, etc., then he would be smug. A low or negative correlation could be taken to mean that the person was either very humble or had a feeling of failure or imperfection. In this case he would be saying, "I know what the relative importance of my values ought to be, but they actually are the opposite."

(*C*) The combination of interpretation of the correlations of 1 with 6, 2 with 6, and 3 with 6 yield an insight into the conceit and sentimentality of the individual. If the correlation of 1 with 6 is positive and high and the correlation of 2 with 6 is low or negative, then the individual certainly is conceited (though we do not know whether he has reason to be or not). If the correlation of 2 with 6 is higher than 1 with 6, we can say he is not very much impressed with himself (though he may have or may not have reason to be). These conclusions do, however, in part depend upon the correlation of 3 with 6, since, if the individual did not subscribe to the order of interests called ideal by the group, his feeling that he was like that would constitute no boast. If, however, the correlation of 1 with 6, and 2 with 6 are both high—say as high as 3 with 6—it shows a naive sentimentality, since it indicates that the individual is calling "typical" what the group calls "ideal." We define sentimentality as theoretical evaluation, hence judgment in terms of desire, not evidence.

(*D*) Different amounts of agreement between 2 and 3, an individual's ideal order and the order he thought the typical had, could mean if the correlation were high and positive a general optimism, *i.e.*, he judges that people have about the relative order of interests that they ought to have. If the correlation was low or negative a critical or even cynical attitude would be suggested, *i.e.*, a feeling on the part of the reporter that whatever else might be true of people, they have values in the wrong order.

(*E*) If there were in any specific situation a negative correlation between 2 and 5, the order of the typical as judged by the subject, and the *actual* order of interest of the group, the recognition of this by a reporter would imply a clear sightedness which many assume, but few possess. An instance of this is: A certain professor was asked to rate the interests as juniors would report them. His rating correlated with our pooled order—58. Another professor rated the interests as he thought they would be rated by juniors. His rating correlated $r=0.89$ with our pooled rating. It seems fair to assert that the second professor knows the student and that the first does not.

(*F*) In the case of the amount of agreement between 3 and 6, the ideal order of interests in the opinion of an individual and the ideal order of interests derived from pooling the ratings of the group, a high positive correlation would suggest that the reporter had values about like the group. Low or negative correlation would suggest at least one important deviation of the individual from his group.

Taking the six orders of interests thus derived: An individual's rating of himself, of his idea of the typical, of his ideal, and these three orders pooled—a study of all the inter-relations on a correlation basis contains many valuable tips for those concerned with character analysis devices.

Another attempt to get the amount of agreement between order of interests taken introspectively and objectively was made with a second group of juniors (71 cases). In the first attempt the interests were rated in order 1 to 34. The correlations may have been lower because the difficulty of the task may have acted as chance error. If this is true, then the low correlations which we have used to prove a tendency of over-rating would change as the difficulty of the experiment was lessened.

In the second attempt we had the group assign values of *A*, *B*, *C*, *D* and *E*,—*A*, for the most important or strongest. The reporters were cautioned to distribute the marks in conformation with the normal

curve in so far as such distribution did not conceal their genuine opinion. The interests were the same as before and order of importance for one's self, the typical junior, and the ideal were obtained.

The following table will show that the disagreement between the

DISTRIBUTION OF JUDGMENTS ON RELATIVE IMPORTANCE OF INTERESTS

Column under *I* is the ideal rating; *S* is the order of importance of interests taken introspectively; *T* is the reports taken objectively. Where too few votes are recorded for comparative purposes none are printed. *A* is strongest interest. *F* is weakest.

	A			*B*			*C*			*D*			*E*			*F*		
	I	*S*	*T*	*I*	*S*	*T*	*I*	*S*	*T*	*I*	*S*	*T*	*I*	*S*	*T*	*I*	*S*	*T*
Parents	23	17	9	8	11	12	2	2	13									
Politics	...	...	...	11	2	1	16	3	11	3	7	8	1	6	10	1	13	3
Dancing	0	1	13	0	7	14	19	15	7	9	5	0	5	3	0			
Children	...	...	...	13	9	5	14	12	10	4	7	6	1	3	7			
Social approval	6	8	18	14	12	12	11	9	3									
Business	...	...	...	20	2	3	9	10	11	3	8	7	0	6	8			
Personal appearance	4	5	7	18	17	22	9	10	4									
Driving a car	...	...	...	3	8	14	13	10	14	9	3	3	5	4	1	3	7	0
Friends	8	5	9	23	21	18	3	5	5									
Sorority	0	2	12	8	13	15	18	4	3	5	4	4						
Art	...	...	...	17	8	4	13	9	11	3	8	12	0	3	4			
Music	0	6	0	19	13	2	13	10	16	1	2	10						
Home	16	15	2	11	8	7	4	6	18	1	3	4						
Athletics	...	...	...	9	5	10	19	15	17	3	3	3	1	4	3			
Studies	9	4	0	19	18	7	3	9	19	2	1	5						
Sleeping	...	...	...	15	4	5	13	14	22	10	6	3	4	6	2			
Social entertainment	...	...	...	3	7	18	23	15	11	5	5	0						
Society	...	...	...	10	4	10	19	10	14	1	6	7	2	7	1			
Health	15	11	1	6	14	8	3	0	15	0	1	6						
Character	26	13	5	7	17	11	1	2	12									
Dress	1	2	6	14	18	18	17	10	9									
Men	0	1	9	8	8	18	15	15	4	6	5	3						
Theatre	...	...	...	3	9	7	18	16	15	7	6	5						
Church	14	2	0	13	6	0	6	14	11	1	6	14	0	3	6			
Eating	...	...	...	1	3	6	21	20	18	6	2	5	4	4	1			
Education	11	4	0	18	20	10	5	8	13	0	0	8						
Reading	5	4	0	23	8	6	5	17	14	0	3	8	0	1	6			
Laughing	...	...	...	9	9	16	20	12	13	3	3	3	0	6	2			
Domestic duties	...	...	...	7	3	0	17	10	7	7	10	12	2	7	7	0	1	7
Magazines	...	...	...	8	6	6	16	10	17	7	12	5	1	2	3			
Travel	5	9	2	14	9	10	14	7	11	0	3	6						
Movies	...	...	...	2	5	15	12	11	9	11	9	3	6	7	0			
Garden	...	...	...	2	2	0	14	3	0	8	0	4	7	10	7	2	18	23
Economy	8	1	0	15	9	1	9	9	11	0	7	9	1	3	7	0	2	6

In 80 instances the self rating was between the *ideal* and the *typical*.

In 39 instances the self rating was not the middle rating. Thus taking the *A* column and the row giving ratings on Parents, 23 Juniors were of the opinion that ideally "pleasing our parents" should be a very strong (*A*) interest. Seventeen recorded it as true of themselves but only 9 thought it was true of the typical junior. Any other set of ratings may be similarly read.

order of interest taken introspectively and objectively still persists. As the introspective order is again nearer the ideal, the data further substantiate the fact of serious tendency of over-rating one's self.

The tendency to place one's self nearer the ideal than the typical, which is shown by the fact that the pooled ratings taken introspectively are nearer the ideal than when taken objectively, is again illustrated by the following data:

Sixty other juniors were asked to rate themselves on the blank presented here, entitled "Self-analysis Test."

SELF-ANALYSIS TEST

Name..

Age..

Years Months

School................................... Grade.........................

City...................................... Date..........................

....

DIRECTIONS: Read each question, then place an X in the column at the head of which is the answer that you think is true

Can you be trusted?	Answers			
	Always	Nearly always	Some-times	Never
1. To do a given task exactly as it was given to you to do?..............				
2. To work faithfully when you work alone as when you are observed?...				
3. To stick to a point when you know you are right?......................				
4. To avoid taking property belonging to others?.....................				
5. To avoid making false claims about yourself?......................				
6. To be fair in an examination?........				
7. To return borrowed property?........				
8. To keep a promise?.................				
9. To repeat a message accurately?......				
10. To be honest in scoring yourself?.....				

They then rated the typical junior on a similar blank. The pooled results deal with the same data; the difference is in the manner of judging, subjective versus objective.

The first number in each pair of the summary sheets is the result of the objective ratings of the same group. The second number of each pair is the result of the introspective ratings.

SUMMARY OF JUDGMENTS ON SELF-ANALYSIS TEST

Can you be trusted?	Answers							
	Always		Nearly always		Some-times		Never	
	Others–	Self	Others–	Self	Others–	Self	Others–	Self
1. To do a given task exactly as it was given to you to do?	1	4	34	50	27	7	0	0
2. To work faithfully when you work alone as when you are observed?	2	23	25	25	34	11	1	0
3. To stick to a point when you know you are right?	15	38	38	19	9	4	0	0
4. To avoid taking property belonging to others?	10	48	35	9	17	3	0	0
5. To avoid making false claims about yourself?	3	38	25	21	33	2	1	1
6. To be fair in an examination?	4	33	36	27	22	1	0	0
7. To return borrowed property?	6	36	24	24	28	1	3	0
8. To keep a promise?	1	24	29	30	32	6	0	0
9. To repeat a message accurately?	3	18	30	24	27	7	2	1
10. To be honest in scoring yourself?	8	39	40	19	13	2	1	1
	53	301			242	44		

Looking at the column headed "Always," which denotes superiority when rating is objective, the total is 53; when introspective, 301. Looking at the column headed "Sometimes," which denotes relative inferiority when the rating is objective, the total is 242—when introspective, only 44.

These data are certainly an illustration of a naive over-rating of one's self, of "putting the best foot forward" proclivity, or of under-estimating one's fellows. As students were specifically told not to

sign their names no conscious desire to cheat anyone else except themselves could have operated greatly.

One other set of data is illuminating in this connection. Using Mendenhall's Moral Character Scale, we get an average correlation of +52 between the relative importance of the traits—in the opinion of the reporter, and the amounts of each trait he thinks he possesses.

In our reporters, a group of graduate students, under-classmen, and professors, there was a clear tendency to speak well of themselves in those virtues considered of greater importance by them, and to rate themselves less highly in traits considered less vital. This positive correlation between the relative importance of traits and the amount of each trait a subject rates himself as possessing, may well be considered a self-defense mechanism whereby a person tends to think well of himself in what he judges important and evens up by under-rating himself in less significant items. Common sense shows that all of us would be readier to admit poor memory or poor handwriting than poor judgment or inferior trustworthiness.[1] This tendency should be figured on when interpreting the self-ratings of an observer. Whether this tendency applies also to the rating of others whom the rater likes, and whether a reverse tendency is present in the judging of those the rater dislikes is unknown.

We had another group of 50 college students rate the qualities used in this test, according to importance, from *A* to *F*. There was some disagreement as to the importance of the qualities which we can neglect here. Each then rated himself from 0 to 20 in each quality.

It is significant to find that these 50 reporters gave themselves an average of

14.75 points on *A* qualities
13.5 points on *B* qualities
11.1 points on *E* qualities
7.0 points on *F* qualities

The tendency to give higher rating to *A* and *B* qualities than are given to *E* and *F* qualities, falls in line with the distribution of correlation between rank order of qualities and amount possessed in each quality.

---

[1] This accounting of the +52 correlation is preferable to a more labored explanation which would contradict normal frequency areas of all of these traits as they actually are in our subjects—assuming then, distributions skewed positively for unimportant traits and skewed negatively for important traits.

These data amply support us in holding that there is a well-marked tendency for a person to overrate himself when he compares himself with others, and even when the introspective judgment is independent of comparison with others, this tendency still persists.

II. The Overlapping of Traits

Another pitfall is involved in analyzed rating schemes. These devices are used to get the rating for the several qualities or traits composing the whole. We find the inter-correlations of ratings for specific traits so high that the general opinion must be present in the ratings given to discrete traits.

Thorndike speaks of this as the spread. When ratings are obtained for traits X, Y and Z of a group the correlations between these ratings show more mutual relationship than could actually exist. These high correlations seem to show a tendency to keep rating the same thing over and over again under different headings. Thus when one gets a correlation of +94 between "quality of voice" and "moral stamina" in a teaching staff,[1] this agreement is taken to mean that the rater is giving his general estimation of teachers when he thinks he rates for voice and again when he thinks he rates for morals. The amount of spread or fusion of common factors found through inter-correlation studies of analyzed ratings of engineers (reported by Thorndike) and of teachers (reported by Knight) is so great that a pretty good case against the usefulness of analyzed ratings can be constructed.

The amount of this spread is a function of the method of rating as well as the inability of judges to rate for specific traits, and therefore it can be partially eliminated. The worst thing about analyzed ratings is not the too high correlations between traits, but the extreme variation of the size of the inter-correlation under different circumstances. This makes it impossible to diagnose the general factor, and to partial it out. To illustrate:

Using the Boice Teachers Rating Card, the inter-correlation on the 40 traits in a teaching staff of over 100 teachers, had a central tendency of +0.5 with variations about like a normal curve. Using a rating card calling for ratings in 10 traits commonly estimated as important for teachers and using the rank order method of rating in a Massachusetts school system (100 teachers), the correlations were all

[1] Knight, F. B.: Qualities Connected with Success in Teaching, *Teachers College Contributions,* in preparation.

too high, many being 0.9. Here every teacher rated every other teacher in the system. In another connection we had exceedingly competent judges rate 100 instructors in a middle-western university for "*savoir faire,*" "research ability," and "ability to teach." The average of these inter-correlations was but +0.32.

Using three different methods for obtaining the same fact, we obtain central tendencies of inter-correlations of 0.3 to 0.9, which means that inter-correlations of traits is a highly variable function of the method of obtaining them.

We conclude here that in analyzed ratings there is a spread of aura present, but in no constant amount. Just how much added information analyzed ratings give over unanalyzed ratings is very uncertain. In using analyzed rating, then, the experimenter must determine the amount of spread indicated by unreasonably high inter-correlations and be careful to interpret his findings in the light of such spread.

INTELLIGENCE TESTS OF FOREIGN CHILDREN

RUDOLPH PINTNER AND RUTH KELLER

Teachers College, Columbia University

In Youngstown, Ohio, the children of the kindergarten, and of the first and second grades of three schools, were classified according to mental age in 1919 and 1920. The test used was a revision of the Binet Test which was prepared and given by members of the Children's Service Bureau of that city, and which showed a correlation of 97 with the Stanford Revision of the Binet Test. It required much less time to give this test than it did the Stanford Revision, and this saving in time was the reason for its use.

A large percentage of the children were of foreign parentage and heard only a foreign language in their homes. It was to discover the influence of this language handicap upon the Binet Test that the following investigation was made.

The nationalities of the children were obtained from cards which the teachers had sent home with the children at the beginning of the school year, requesting such information as the child's name, address, age, birth date, father's name, and nationality.

The child's chronological age was then verified by the teacher from birth certificate, or church records where possible. The records of only those children whose ages were thus verified were used in the preparation of this article.

The records of Jewish children were eliminated from the totals unless the cards definitely stated whether they were of American or foreign parentage. The English speaking group includes American white, Negro, English, Canadian, Scotch, Irish, and Welsh. The foreign group is predominately Italian and Slavish, but includes also the following nationalities: German, Greek, Hungarian, Polish, Finnish, Croatian, Austrian, French, Swedish, Syrian, Gypsy, Lithuanian, Roumanian, Spanish, Russian, and Indian. No attempt was made to distinguish between race and nationality, but they were noted just as they were written on the cards.

Although the number of cases was not sufficient to make the results very significant, the average Intelligence Quotient for each nationality is listed below for whatever interest it might contain.

The average for all the 79 Jewish children irrespective of whether they spoke English or a foreign language was 95.

English speaking			Foreign speaking		
Number of cases	Nationality	IQ	Number of cases	Nationality	IQ
249	American (white)	95	313	Italian	84
71	American (colored)	88	130	Slavish	85
24	English	97	99	Hungarian	89
3	Canadian	89	37	German	91
8	Scotch	88	18	Roumanian	97
5	Irish	92	12	Greek	83
7	Welsh	93	11	Polish	85
			10	Russian	89
			7	Lithuanian	87
			5	Croatian	86
			4	Syrian	80
			4	Gypsy	74
			4	Finnish	94
			3	Austrian	94
			3	Swedish	104
			2	Spanish	93
			1	Indian	93
			1	French	125

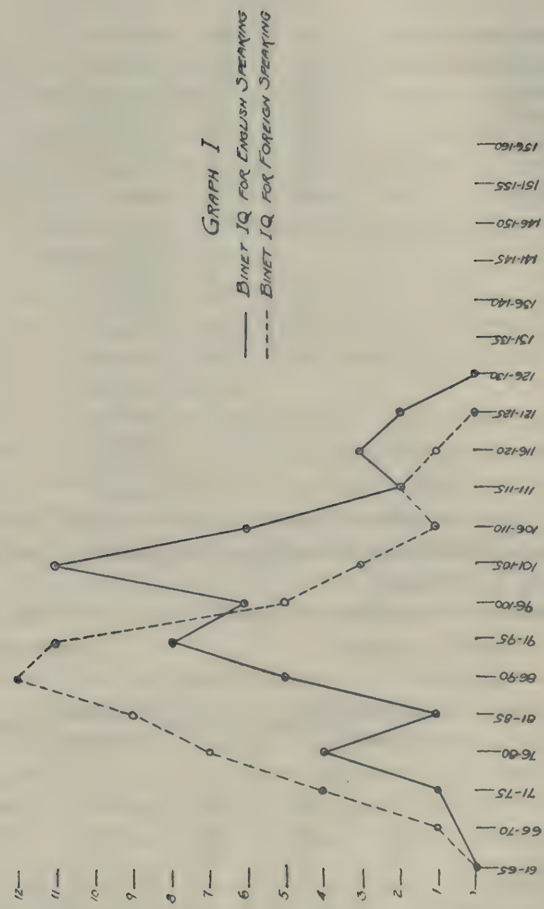
The average and median Intelligence Quotient for each of the three schools is as follows:

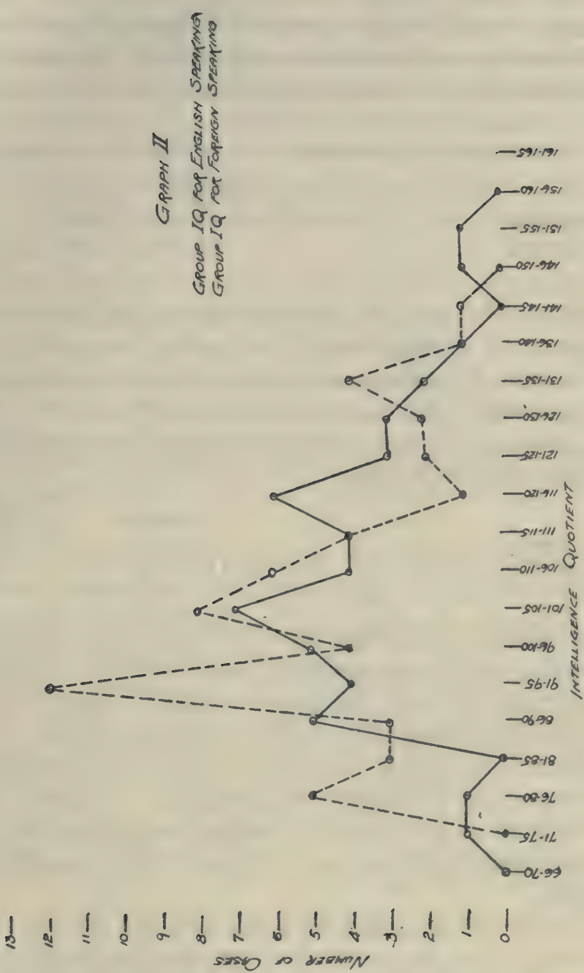
	English speaking			Foreign speaking		
	Number of cases	Average	Median	Number of cases	Average	Median
School I	145	96	97	230	86	86
School II	172	90	93	261	83	83
School III	50	91	91	183	87	86

From the above table, it will be seen that both the average and the median for the English speaking children of all three schools are significantly higher than those of the foreign speaking children.

The average and median Intelligence Quotients for all the English and all the foreign speaking children in the three schools are:

	Number of Cases	Average	Median
English	367	92	94
Foreign	674	84	85


GRAPH I
—— BINET IQ FOR ENGLISH SPEAKING
---- BINET IQ FOR FOREIGN SPEAKING
61-65
66-70
71-75
76-80
81-85
86-90
91-95
96-100
101-105
106-110
111-115
116-120
121-125
126-130
131-135
136-140
141-145
146-150
151-155
156-160
1
2
3
4
5
6
7
8
9
10
11
12


GRAPH II
GROUP IQ FOR ENGLISH SPEAKING
GROUP IQ FOR FOREIGN SPEAKING
INTELLIGENCE QUOTIENT
NUMBER OF CASES

It is obvious, therefore, that the foreign child rates decidedly lower on the Binet Scale, whether because of actual lower intelligence, or because of a language handicap.

The Pintner Non-language Test and the Binet Test.—The Pintner Non-language Group Test was given to the second grades in School No. I, and following are the results compared with the results obtained by the Children's Service Bureau Revision of the Binet Test.

Average Intelligence Quotients for the Binet and Group Tests

	Number of Cases	Binet IQ	Group IQ
English speaking....................	49	99	109
Foreign speaking....................	56	89	103

The difference between the average IQ obtained by the Binet Test and that obtained by the Group Test which requires the use of a minimum amount of language was for the English speaking 10 and for the Foreign speaking 14 in favor of the Group Test.

The number and per cent of cases in which the Group IQ is higher than the Binet IQ is as follows:

	Number	Per Cent of Total
English speaking......................	36	73
Foreign speaking......................	46	82

In the accompanying graphs are seen the number of cases of both English speaking and foreign speaking found with IQ's within a certain range. In Graph I are the comparative IQ's for the Binet Test, and in Graph II are the comparative IQ's for the Group Test.

Performance Tests and the Binet Test.—From the office files all cases were taken to whom the Stanford Revision of the Binet Test and a series of at least three performance tests had been given, and the results of the English speaking group were compared with those of the foreign speaking group.

The series of performance tests given included all, or at least three of the following: The Pintner Cube Test, the Form Board, the Witmer Cylinders Test, Healy Construction Puzzle A, and the Mare and Foal Test. The subjects ranged in age from very young children to adults. They were court cases, Children's Home cases, medical cases, in fact, all the usual types of cases found in a psychological clinic with a large percentage of foreign cases.

The necessary data concerning nationality, age, and birth date,

were obtained from parents if they were present, from court records where possible, or from some responsible person who accompanied the child, or from the child himself if his age and degree of intelligence seemed to warrant that he was a reliable source of information.

The median mental age of the series of performance tests was used for the comparison. The Binet mental age was correlated with the performance age of both the English speaking and the foreign speaking, and the correlations are as follows: English speaking, 64; foreign speaking 48.

The average amounts of difference between the mental age and the performance age expressed in months are: for the English speaking 6; for the foreign speaking 16.

The number and per cent of cases where the performance age is higher than the mental age are:

	Number of Cases	Per Cent of Total
English speaking	45	52
Foreign speaking	95	75

The accompanying table compares the average mental age with the average performance age for groups of each six months of chronological age.

With the English speaking group, the average performance age is higher in ten groups, lower in seven groups, and equal in two groups. With the foreign speaking, the average preformance age is higher in twenty groups and lower in two groups.

Summary and Conclusions.—A revision of the Binet Test given to children in three schools in which a large majority were foreign speaking, gave the following results: Average IQ for the English speaking 92; average IQ for the foreign speaking 84.

The Pintner Group Test given to one group of these children showed a much higher IQ both for the English speaking and the foreign speaking, but for the foreign speaking, the difference between the results of the two tests was greater and in favor of the foreign group.

In comparing the results obtained from a group of cases given the Stanford Revision of the Binet Test, and a series of Performance Tests, we find the correlation between the tests considerably better for the English speaking group than for the foreign speaking, and there were twenty-three per cent more cases of foreign speaking children than of English speaking where the performance age was higher than the mental age.

English speaking Chronological age.[1]	5–1 to 5–5	5–6 to 6–0	6–1 to 6–5	6–6 to 7–0	7–1 to 7–5	7–6 to 8–0	8–1 to 8–5	8–6 to 9–0	9–1 to 9–5	9–6 to 10–0	10–1 to 10–5	10–6 to 11–0	11–1 to 11–5	11–6 to 12–0	12–1 to 12–5	12–6 to 13–0	13–1 to 13–5	13–6 to 14–0	14–1 to 14–5	14–6 to 15–0	15–1 to 15–5	15–6 to 16–0	Adult
Average MA (months)........	56	...	...	...	92	86	100	94	93	113	116	106	121	146	113	122	136	125	134	119	121	....	134
Average PA (months)........	78	...	...	...	94	85	84	94	97	114	103	119	100	127	107	122	154	145	127	142	164	....	142
Number of cases...	2	0	0	0	2	3	2	9	4	5	4	5	2	5	6	6	3	3	6	9	2	0	9

Foreign Chronological age.	5–1 to 5–5	5–6 to 6–0	6–1 to 6–5	6–6 to 7–0	7–1 to 7–5	7–6 to 8–0	8–1 to 8–5	8–6 to 9–0	9–1 to 9–5	9–6 to 10–0	10–1 to 10–5	10–6 to 11–0	11–1 to 11–5	11–6 to 12–0	12–1 to 12–5	12–6 to 13–0	13–1 to 13–5	13–6 to 14–0	14–1 to 14–5	14–6 to 15–0	15–1 to 15–5	15–6 to 16–0	Adult 7
Average MA (months)........	...	52	52	54	57	64	67	71	82	71	92	93	102	103	100	101	102	106	99	104	105	109	110
Average PA (months)........	...	80	72	82	84	82	78	83	91	95	120	104	100	108	113	130	124	117	98	118	134	136	129
Number of cases..	0	1	2	5	2	3	3	3	8	7	3	8	9	11	10	10	10	7	6	5	1	6	7

From these results, we may conclude that children who hear a foreign language at home, test lower as a rule when given the revisions of the Binet Test than when given tests which require a minimum knowledge of English. And that when classified according to mental age, those children who hear a foreign language in their homes may suffer a serious handicap when tested only by the revisions of the Binet Test.

THE CORRELATIONS OF ACHIEVEMENT IN SCHOOL SUBJECTS WITH INTELLIGENCE TESTS AND OTHER VARIABLES (CONTINUED)

ARTHUR I. GATES

Teachers College, Columbia University

PART III. RESULTS FOR GRADES IV TO VIII

1. *The Relation between Verbalness and the Magnitude of Correlations with the Criterion.*—Most of the tests used in grades IV and up are verbal or mixed. In order to get a range of material from the most non-verbal to the most verbal, the various exercises in the tests were arranged, according to the judgments of a group of workers familiar with test construction, in a scale with units from zero, extremely non-verbal, to 7.0 extremely verbal.[1] The list is too long to print. At the non-verbal extreme are such tests as Dearborn II, Nos. 7, 3, 9, Myers 3, 4, Haggerty A2, No. 3, etc. and at the other extreme, N. I. T. –B, No. 3; N. I. T.–A, No 2; Otis Advanced No. 9; Dearborn II, No. 4.

The scale was divided into four sections each of which included sufficient tests to represent about one hour of working time. The scores for the exercises in each section were summated and the coëfficients with the criteria obtained from the records for grades IV, V and VI as one group and for grades VII and VIII as another. The correlations have, of course, no validity except for comparisons within the same group.

The results appear in Tables III and IV. The verbal materials yield a higher correlation with mental age than the non-verbal, although the difference between the third and fourth steps is very small. Competent judges place the Stanford test in group 3. Its correlation is slightly higher, however, with group 4 which is more verbal.

In both groups of subjects, the second step on the verbal scale gives the highest correlation with arithmetic. Most arithmetical problems are judged to be of about that degree of verbalness.

If the tests become more verbal than arithmetic in general is judged to be, the correlation drops, although the drop may really

[1] This scale which was originally constructed by Dr. John Herring is described in detail in a thesis, as yet unpublished, in the library at Teachers College. Dr. Herring's scale was used as a framework to which additional exercises in our tests were affixed.

be due to the greater identity of content than to degree of verbalness in general.

For both groups, the higher the material on the verbal scale, the higher the correlation with spelling. The same is true of reading, which shows the greatest increase in correlation as the material becomes more verbal. Competent judges place both of these subjects high on the verbal scale.

The composite of achievement, which is weighted heavily by the verbal subjects, shows an increasing correlation as the tests became more verbal. The correlations with the composite are higher than those with single subjects partly because the composite is a more reliable measure and doubtless partly because each of the component subjects has an independent partial correlation with the intelligence tests.

These results for the upper grades agree in all essentials with the findings for grade III.

Table III.—Correlations for a Group Composed of Pupils of Grades IV, V and VI. n = 63

	1 Mental age	2 Arithmetic	3 Spelling	4 Reading	5 Composite achievement
1. Most non-verbal	0.40	0.39	0.18	0.32	0.46
2. Somewhat verbal	0.49	0.68	0.38	0.39	0.62
3. More verbal	0.68	0.59	0.46	0.65	0.75
4. Most verbal	0.73	0.49	0.49	0.76	0.79

Table IV.—Correlations for a Group Composed of Pupils of Grades VII and VIII. n = 42

	1 Arithmetic	2 Spelling	3 Reading	4 Composite achievement
1. Most non-verbal	0.17	0.09	0.03	0.10
2. Somewhat verbal	0.49	0.29	0.49	0.56
3. More verbal	0.29	0.38	0.53	0.55
4. Most verbal	0.24	0.44	0.67	0.60

2. *The Relation between the Magnitude of the Correlations and the Length of the Test.*—In Parts I and II, high positive correlations were found between the length of the test and the magnitude of the correlations with the criterion. In the case of grades IV to VIII, it is less easy to make this comparison for the reason that the tests also differ in the degree of verbal material which they contain.

TABLE V.—SHOWING THE LENGTH IN MINUTES, THE DEGREE OF VERBALNESS ON A SCALE IN WHICH 1.0 IS VERY NON-VERBAL, AND 4.0 VERY VERBAL AND THE AVERAGE r WITH ACHIEVEMENT FOR VARIOUS TESTS

	1 Time (minutes)	2 Verbalness	3 Mean r with achievement
Dearborn Total	80	1.8	0.47
Otis advanced	47	3.0	0.63
Dearborn 5	45	1.8	0.43
Dearborn 4	35	1.8	0.38
National Total	33	2.6	0.63
Thorndike-McCall	30	3.6	0.48
Terman Groups	27	3.2	0.55
Haggerty, Delta 2	21	2.3	0.52
National A	17	2.6	0.56
Illinois	17	2.6	0.48
National B	16	2.6	0.66
Myers	15	1.0	0.12
Holley	12	3.0	0.43

Following are the simple correlations of the columns:

1. Achievement with verbalness 0.67
2. Achievement with time . . . . –0.04
3. Time with verbalness . . . . –0.28

Since the length of tests and verbalness are negatively correlated, the simple coëfficients do not display clearly the influence of either.

By use of the regression equation, the proper independent weight of time and verbalness in determining the correlation with achievement are:

4. Weight of verbalness 1.00
5. Weight of time (β) 0.224

The independent correlations of time and verbalness can best be

shown by partial correlation of each with the criterion when the other is held constant. The partial correlations, first order, are:

6. Partial r, criterion with verbalness (time constant).... 0.69
7. Partial r, criterion with time (verbalness constant).... 0.21

Both time and the degree of verbalness are determining factors, but verbalness alone yields a higher correlation, with achievement than time alone. That is to say for purposes of predicting school achievement it is better to have a very verbal test than a very long test, if one cannot have both. The best thing to have is a long verbal test, as the following multiple correlation shows.

8. Multiple r, achievement with (verbal + time)......... 0.73

This figure should be compared with 6 and 7 above.

3. *The Validity of the Decidedly Non-verbal Tests; Grade Differences.* Table VI gives the correlations for each test for grades IV–VIII separately with the means of the grade columns. Column 14 gives the mean correlation of each test with other group tests. In computing the mean inter-correlations, the correlations of parts of the Dearborn and National Tests with the total, and the total with the parts have been omitted.

The Myers Test is wholly non-verbal, the Dearborn contains a great deal of non-verbal material, but the other tests are largely verbal.

Study of Table VI discloses the fact that the Myers tests shows small inter-correlations with the verbal tests. The mean inter-correlation with all other tests is 0.21 as compared to 0.47, the mean of the mean inter-correlation of all other tests. The correlations of the Myers with the various measures of achievement are very low compared to others. This is in line with the expectations set up by the foregoing sections. The non-verbal material doubtless provides a useful measure of some human traits, but in these grades at least, not with achievement in school subjects and other functions which are largely verbal.

It should be noted that the Myers test agrees with other criteria much more closely in grades IV, V and VI than in VII and VIII. The Myer's correlations with Mental Age also drops rapidly as the grade becomes higher.

The Myers test correlates more closely with the Dearborn Examinations than with others. This, we are quite certain, is due to the fact that the latter test is more non-verbal than any others. The Dearborn contains considerable verbal material and agrees more closely with other group tests than does the Myers. The mean correlation of

Table VI.—Correlations Grades IV–VIII

Grade	1 Myers	2 Dearborn 4	3 Dearborn 5	4 Dearborn 4 and 5	5 National Intelligence Test A	6 National Intelligence Test B	7 National Intelligence Test A and B	8 Holley Vocabulary	9 Otis Advanced	10 Haggerty Delta 2	11 Illinois	12 Terman Group	13 Thorndike-McCall	14 Mean Intercorrelation	15 Stanford Mental Age	16 Reading Comprehension	17 Reading Rate	18 Arithmetic	19 Spelling	20 Composite Achievement
Correlations of Myers Mental Measure with																				
IV	......	0.45	0.71	0.56	0.25	0.23	0.46	0.15	......	0.20	0.29	....	0.08	0.29	0.41	0.17	0.10	0.27	0.09	0.22
V	......	0.76	0.47	0.55	0.30	0.25	0.52	0.21	0.40	0.16	0.23	....	0.26	0.29	0.29	0.15	0.12	0.16	0.18	0.17
VI	......	0.42	0.51	0.54	0.44	0.28	0.40	0.30	0.50	0.40	0.46	....	0.18	0.34	0.15	0.26	0.30	0.21	0.21	0.29
VII	......	0.52	0.35	0.40	−0.10	0.00	−0.05	0.30	0.29	0.19	−0.24	0.15	0.13	0.14	....	0.22	−0.03	0.08	0.08	0.08
VIII	......	0.46	0.39	0.52	−0.30	−0.38	−0.40	−0.18	−0.37	−0.06	−0.11	0.07	−0.30	−0.03	....	−0.26	−0.38	0.00	−0.03	−0.15
Mean.......	......	0.52	0.49	0.51	0.12	0.08	0.15	0.16	0.21	0.16	0.13	0.11	0.07	0.21	0.28	0.11	0.02	0.14	0.11	0.12
SD.........	......	0.12	0.13	0.06	0.27	0.25	0.33	0.18	0.34	0.15	0.29	0.04	0.19	0.14	0.11	0.16	0.23	0.10	0.08	0.15
Correlations, Dearborn Examination 4 with																				
IV	0.45	......	0.93	0.95	0.44	0.33	0.52	0.39	......	0.56	0.48	....	0.33	0.43	0.28	0.37	0.14	0.37	0.39	0.44
V	0.76	......	0.73	0.90	0.50	0.34	0.45	0.47	0.56	0.42	0.50	....	0.53	0.48	0.59	0.42	0.28	0.36	0.14	0.45
VI	0.42	......	0.78	0.89	0.64	0.47	0.59	0.55	0.66	0.70	0.59	....	0.65	0.58	0.68	0.65	0.59	0.35	0.30	0.60
VII	0.52	......	0.61	0.88	0.27	0.27	0.38	0.34	0.53	0.48	0.32	0.46	0.59	0.41	....	0.52	0.55	0.22	0.21	0.30
VIII	0.46	......	0.53	0.77	0.08	−0.06	0.06	0.00	0.12	0.27	0.05	0.43	0.06	0.15	....	0.12	−0.05	0.15	0.04	0.09
Mean.......	0.52	......	0.72	0.88	0.39	0.27	0.40	0.35	0.47	0.49	0.39	0.45	0.43	0.41	0.52	0.42	0.30	0.29	0.22	0.38
SD.........	0.12	......	0.14	0.06	0.19	0.17	0.18	0.19	0.21	0.14	0.19	0.02	0.21	0.14	0.17	0.18	0.24	0.09	0.12	0.17

TABLE VI.—(*Continued*)

Grade	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	Myers	Dearborn 4	Dearborn	Dearborn 4 and 5	National Intelligence Test A	National Intelligence Test B	National Intelligence Test A and B	Holley Vocabulary	Otis Advanced	Haggerty Delta 2	Illinois	Terman Group	Thorndike-McCall	Mean Intercorrelation	Stanford Mental Age	Reading Comprehension	Reading Rate	Arithmetic	Spelling	Composite Achievement
Correlations, Dearborn Examination 5 with																				
IV	0.71	0.93	....	0.97	0.45	0.41	0.73	0.26	......	0.72	0.69	....	0.46	0.57	0.48	0.49	0.34	0.49	0.57	0.64
V	0.47	0.73	....	0.98	0.52	0.41	0.76	0.40	0.62	0.45	0.60	....	0.63	0.48	0.44	0.51	0.26	0.28	0.19	0.47
VI	0.51	0.78	....	0.88	0.55	0.45	0.57	0.46	0.70	0.55	0.52	....	0.52	0.51	0.56	0.66	0.45	0.28	0.15	0.49
VII	0.35	0.61	....	0.86	0.21	0.43	0.40	0.29	0.37	0.32	0.17	0.59	0.34	0.34	....	0.32	0.29	0.18	0.13	0.39
VIII	0.39	0.53	....	0.69	0.22	0.12	0.30	0.00	0.22	0.36	0.40	0.50	0.06	0.24	....	0.28	0.14	0.21	0.02	0.18
Mean.......	0.49	0.72	....	0.88	0.39	0.36	0.55	0.28	0.48	0.48	0.48	0.55	0.40	0.43	0.49	0.45	0.30	0.31	0.21	0.43
SD.........	0.13	0.14	....	0.10	0.15	0.12	0.18	0.19	0.19	0.14	0.18	0.05	0.21	0.12	0.05	0.14	0.10	0.15	0.19	0.15
Correlations, Dearborn Total with																				
IV	0.56	0.95	0.97	....	0.42	0.41	0.54	0.36	......	0.68	0.59	....	0.51	0.51	0.50	0.49	0.39	0.60	0.44	0.61
V	0.55	0.90	0.98	....	0.59	0.51	0.62	0.47	0.66	0.48	0.58	....	0.65	0.51	0.55	0.58	0.33	0.42	0.18	0.55
VI	0.54	0.89	0.88	....	0.56	0.54	0.57	0.55	0.74	0.65	0.60	....	0.68	0.57	0.68	0.70	0.46	0.33	0.27	0.53
VII	0.40	0.88	0.86	....	0.31	0.39	0.44	0.36	0.48	0.46	0.28	0.57	0.48	0.40	....	0.53	0.44	0.29	0.19	0.44
VIII	0.52	0.77	0.69	....	0.16	0.09	0.21	0.00	0.24	0.36	0.21	0.44	0.12	0.22	....	0.31	0.12	0.21	0.09	0.20
Mean.......	0.51	0.88	0.88	....	0.41	0.39	0.48	0.35	0.53	0.53	0.45	0.51	0.49	0.44	0.58	0.52	0.35	0.37	0.23	0.47
SD.........	0.06	0.06	0.10	....	0.15	0.15	0.15	0.20	0.17	0.12	0.17	0.07	0.20	0.12	0.09	0.13	0.12	0.10	0.12	0.14

Correlations of National, Form A with

IV	0.25	0.44	0.45	0.42	......	0.66	0.79	0.36	......	0.59	0.59	....	0.51	0.48	0.38	0.63	0.49	0.35	0.52	0.54
V	0.30	0.50	0.52	0.59	......	0.61	0.87	0.53	0.61	0.49	0.76	....	0.80	0.53	0.50	0.70	0.54	0.21	0.35	0.48
VI	0.44	0.64	0.55	0.56	......	0.68	0.90	0.47	0.71	0.55	0.74	....	0.46	0.54	0.53	0.55	0.56	0.20	0.39	0.59
VII	−0.10	0.27	0.21	0.31	......	0.74	0.92	0.26	0.55	0.63	0.79	0.48	0.37	0.43	....	0.57	0.61	0.26	0.29	0.58
VIII	−0.30	0.08	0.22	0.16	......	0.86	0.97	0.16	0.58	0.50	0.73	0.44	0.37	0.40	....	0.50	0.75	0.29	0.40	0.63
Mean.......	0.12	0.39	0.39	0.41	......	0.70	0.89	0.36	0.61	0.55	0.72	0.46	0.51	0.48	0.47	0.59	0.59	0.26	0.39	0.56
SD.........	0.27	0.19	0.15	0.15	......	0.08	0.06	0.13	0.06	0.05	0.07	0.02	0.16	0.05	0.06	0.07	0.09	0.05	0.08	0.05

Correlations of National, Form B with

IV	0.23	0.33	0.41	0.41	0.66	......	0.77	0.41	......	0.58	0.62	....	0.56	0.47	0.40	0.57	0.40	0.25	0.60	0.67
V	0.25	0.34	0.41	0.51	0.61	......	0.85	0.31	0.66	0.53	0.47	....	0.41	0.47	0.45	0.66	0.45	0.27	0.58	0.54
VI	0.28	0.47	0.45	0.54	0.68	......	0.89	0.50	0.67	0.46	0.68	....	0.61	0.54	0.50	0.51	0.57	0.47	0.43	0.70
VII	0.00	0.27	0.43	0.39	0.74	......	0.90	0.17	0.55	0.51	0.77	0.51	0.34	0.46	....	0.59	0.60	0.45	0.41	0.68
VIII	−0.38	−0.06	0.12	0.09	0.86	......	0.96	0.20	0.56	0.61	0.62	0.47	0.39	0.39	....	0.48	0.69	0.36	0.49	0.69
Mean.......	0.08	0.27	0.36	0.39	0.70	......	0.87	0.32	0.61	0.54	0.63	0.49	0.46	0.47	0.45	0.56	0.54	0.36	0.50	0.66
SD.........	0.25	0.17	0.12	0.15	0.08	......	0.06	0.12	0.06	0.05	0.10	0.02	0.10	0.05	0.04	0.06	0.10	0.09	0.08	0.06

Correlations, National, Forms A and B (total) with

IV	0.46	0.52	0.73	0.54	0.79	0.77	......	0.39	......	0.63	0.64	....	0.56	0.53	0.46	0.64	0.45	0.32	0.55	0.65
V	0.32	0.45	0.76	0.62	0.87	0.85	......	0.48	0.73	0.55	0.61	....	0.53	0.54	0.52	0.72	0.55	0.29	0.48	0.56
VI	0.40	0.59	0.57	0.57	0.90	0.89	......	0.53	0.72	0.58	0.68	....	0.58	0.56	0.58	0.63	0.60	0.33	0.43	0.63
VII	−0.05	0.38	0.40	0.44	0.92	0.90	......	0.27	0.58	0.66	0.80	0.56	0.42	0.48	....	0.64	0.65	0.36	0.36	0.66
VIII	−0.40	0.06	0.30	0.21	0.97	0.96	......	0.22	0.60	0.60	0.66	0.50	0.44	0.40	....	0.58	0.68	0.31	0.46	0.66
Mean.......	0.15	0.40	0.55	0.48	0.89	0.87	......	0.38	0.66	0.60	0.68	0.53	0.51	0.50	0.51	0.64	0.59	0.32	0.46	0.63
SD.........	0.33	0.18	0.18	0.15	0.06	0.06	......	0.12	0.05	0.04	0.07	0.03	0.06	0.06	0.05	0.04	0.08	0.02	0.06	0.04

TABLE VI.—(*Contined*)

Grade	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	Myers	Dearborn	Dearbon 5	Dearborn 4 and 5	National Intelligence Test A	National Intelligence Test B	National Intelligence Test A and B	Holley Vocabulary	Otis Advanced	Haggerty Delta 2	Illinois	Terman Group	Thornlike-McCall	Mean Intercorrelation	Stanford Mental Age	Reading Comprehension	Reading Rate	Arithmetic	Spelling	Composite Achievement
Correlations, Holley's Vocabulary with																				
IV	0.15	0.39	0.26	0.36	0.36	0.41	0.39	......	......	0.39	0.38	....	0.61	0.40	0.38	0.73	0.73	0.04	0.44	0.42
V	0.21	0.47	0.40	0.47	0.53	0.31	0.48	......	0.43	0.42	0.33	....	0.49	0.40	0.38	0.56	9.41	0.20	0.20	0.44
VI	0.30	0.55	0.46	0.55	0.47	0.50	0.53	......	0.70	0.44	0.46	....	0.58	0.49	0.51	0.65	0.63	0.08	0.45	0.49
VII	0.30	0.34	0.29	0.36	0.26	0.17	0.27	......	0.54	0.38	0.29	0.52	0.43	0.36	....	0.50	0.57	0.12	0.39	0.43
VIII	−0.18	0.00	0.00	0.00	0.16	0.20	0.22	......	0.33	0.12	0.24	0.28	0.54	0.22	....	0.48	0.31	0.24	0.40	0.39
Mean.......	0.16	0.35	0.28	0.35	0.36	0.32	0.38	......	0.50	0.35	0.34	0.40	0.54	0.37	0.42	0.60	0.55	0.14	0.38	0.43
SD.........	0.18	0.19	0.19	0.20	0.13	0.12	0.12	......	0.14	0.12	0.08	0.12	0.06	0.09	0.06	0.10	0.15	0.07	0.09	0.03
Correlations, Otis Advanced with																				
V	0.40	0.56	0.62	0.66	0.61	0.66	0.73	0.43	......	0.44	0.60	....	0.75	0.53	0.55	0.73	0.49	0.10	0.23	0.50
VI	0.50	0.66	0.70	0.74	0.71	0.67	0.72	0.70	......	0.74	0.59	....	0.72	0.64	0.66	0.61	0.64	0.40	0.41	0.65
VII	0.29	0.53	0.37	0.48	0.55	0.55	0.58	0.54	......	0.58	0.57	0.70	0.51	0.53	....	0.77	0.70	0.26	0.40	0.66
VIII	−0.37	0.12	0.22	0.24	0.58	0.56	0.60	0.33	......	0.78	0.41	0.71	0.47	0.43	....	0.60	0.56	0.38	0.50	0.69
Mean.......	0.21	0.47	0.48	0.53	0.61	0.61	0.66	0.50	......	0.64	0.54	0.71	0.61	0.53	0.61	0.68	0.60	0.29	0.39	0.63
SD.........	0.34	0.21	0.19	0.17	0.06	0.06	0.06	0.14	......	0.14	0.08	0.01	0.12	0.07	0.06	0.07	0.08	0.12	0.10	0.07

Correlations, Haggerty Delta 2 with

IV	0.20	0.56	0.72	0.68	0.59	0.58	0.63	0.39	......	......	0.47	....	0.72	0.49	0.24	0.62	0.36	0.12	0.49	0.46
V	0.16	0.42	0.45	0.48	0.49	0.53	0.55	0.42	0.44	......	0.57	....	0.46	0.45	0.57	0.66	0.67	0.09	0.26	0.49
VI	0.40	0.70	0.55	0.65	0.55	0.46	0.58	0.44	0.74	......	0.44	....	0.62	0.53	0.64	0.60	0.49	0.35	0.30	0.55
VII	0.19	0.48	0.32	0.46	0.63	0.51	0.66	0.38	0.58	......	0.54	0.68	0.48	0.51	....	0.69	0.64	0.30	0.43	0.61
VIII	−0.06	0.27	0.36	0.36	0.50	0.61	0.60	0.12	0.78	......	0.52	0.71	0.51	0.43	....	0.59	0.36	0.27	0.41	0.48
Mean.......	0.16	0.49	0.48	0.53	0.55	0.54	0.60	0.35	0.64	......	0.51	0.70	0.56	0.48	0.48	0.63	0.50	0.23	0.38	0.52
SD.........	0.15	0.14	0.14	0.12	0.05	0.05	0.04	0.12	0.14	......	0.05	0.02	0.10	0.04	0.17	0.04	0.13	0.10	0.09	0.06

Correlations, Illinois with

IV	0.29	0.48	0.69	0.59	0.59	0.62	0.64	0.38	......	0.47	......	....	0.34	0.48	0.28	0.59	0.55	0.22	0.46	0.45
V	0.23	0.50	0.60	0.58	0.76	0.47	0.61	0.33	0.60	0.57	......	....	0.66	0.50	0.50	0.65	0.52	0.17	0.26	0.46
VI	0.46	0.59	0.52	0.60	0.74	0.68	0.68	0.46	0.59	0.44	......	....	0.58	0.54	0.56	0.55	0.42	0.41	0.40	0.50
VII	−0.24	0.32	0.17	0.28	0.79	0.77	0.86	0.29	0.57	0.54	......	0.63	0.59	0.47	....	0.67	0.70	0.28	0.32	0.52
VIII	−0.11	0.05	0.40	0.21	0.73	0.62	0.66	0.24	0.41	0.52	......	0.43	0.49	0.40	....	0.60	0.45	0.19	0.35	0.49
Mean.......	0.13	0.39	0.48	0.45	0.72	0.63	0.68	0.34	0.54	0.51	......	0.53	0.53	0.48	0.45	0.61	0.53	0.27	0.36	0.48
SD.........	0.29	0.19	0.18	0.17	0.07	0.10	0.07	0.08	0.08	0.05	......	0.10	0.11	0.05	0.12	0.04	0.10	0.08	0.07	0.03

Correlations, Terman Group with

VIII	0.15	0.46	0.59	0.57	0.48	0.51	0.56	0.52	0.70	0.68	0.63	....	0.58	0.51	....	0.61	0.47	0.18	0.41	0.52
IX	0.07	0.43	0.50	0.44	0.44	0.47	0.50	0.28	0.71	0.71	0.43	....	0.54	0.47	....	0.66	0.55	0.23	0.39	0.58
Mean.......	0.11	0.45	0.55	0.51	0.46	0.49	0.53	0.40	0.71	0.70	0.53	....	0.56	0.49	....	0.64	0.51	0.21	0.40	0.55
SD.........	0.04	0.02	0.05	0.07	0.02	0.02	0.03	0.12	0.01	0.02	0.10	....	0.02	0.02	....	0.03	0.04	0.03	0.01	0.03

TABLE VI.—*Continued.*

Grade	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	Myers	Dearborn 4	Dearborn 5	Dearborn 4 and 5	National Intelligence Test A	National Intelligence Test B	National Intelligence Test A and B	Holley Vocabulary	Otis Advanced	Haggerty Delta 2	Illinois	Terman Group	Thorndike-McCall	Mean Intercorrelation	Stanford Mental Age	Reading Comprehension	Reading Rate	Arithmetic	Spelling	Composite Achievement
Correlations, Thorndike-McCall with																				
IV	0.08	0.33	0.46	0.51	0.51	0.56	0.56	0.61	......	0.72	0.34	....	......	0.44	0.35	0.76	0.43	0.10	0.23	0.47
V	0.26	0.53	0.63	0.65	0.86	0.41	0.53	0.49	0.75	0.46	0.66	....	......	0.52	0.62	0.76	0.45	0.02	0.29	0.46
VI	0.18	0.65	0.52	0.68	0.46	0.61	0.58	0.58	0.72	0.62	0.58	....	......	0.54	0.75	0.64	0.42	0.31	0.32	0.49
VII	0.13	0.59	0.34	0.48	0.37	0.34	0.42	0.43	0.51	0.48	0.59	0.58	......	0.43	....	0.52	0.50	0.12	0.39	0.47
VIII	−0.30	0.06	0.06	0.12	0.37	0.39	0.44	0.54	0.47	0.51	0.49	0.54	......	0.35	....	0.81	0.50	0.09	0.40	0.50
Mean.......	0.07	0.43	0.40	0.49	0.51	0.46	0.51	0.54	0.61	0.56	0.53	0.56	......	0.46	0.57	0.73	0.52	0.13	0.33	0.48
SD.........	0.19	0.21	0.21	0.20	0.16	0.10	0.06	0.06	0.12	0.10	0.11	0.02	......	0.07	0.17	0.12	0.14	0.10	0.06	0.01
Correlations of Stanford Mental Age with																				
IV	0.41	0.28	0.48	0.50	0.38	0.40	0.46	0.38	......	0.24	0.28	....	0.35	0.38	....	0.36	0.23	0.35	0.11	0.43
V	0.29	0.59	0.44	0.55	0.50	0.45	0.52	0.38	0.55	0.57	0.50	....	0.62	0.52	....	0.41	0.56	0.25	0.37	0.51
VI	0.15	0.68	0.56	0.68	0.53	0.50	0.58	0.51	0.66	0.64	0.56	....	0.75	0.60	....	0.69	0.60	0.30	0.45	0.67
Mean.......	0.28	0.52	0.49	0.58	0.47	0.45	0.51	0.42	0.61	0.48	0.45	....	0.57	0.50	....	0.49	0.46	0.30	0.31	0.54
SD.........	0.11	0.17	0.05	0.09	0.05	0.04	0.05	0.06	0.06	0.17	0.12	....	0.17	0.09	....	0.15	0.17	0.04	0.15	0.10

Dearborn Total with Myers is 0.51, while the mean correlations Dearborn with all tests is 0.44. It should be noted also that correlations of Dearborn with the verbal tests become distinctly lower after grade VI, whereas the correlation or Dearborn with Myers shows no marked change. The correlations of Dearborn Total and the composite of achievement show a steady decline from 0.61 for grade IV to 0.20 for grade VIII.

So far as these data are concerned, there is a clear indication that the non-verbal material becomes less valid for the prediction of success in school subjects as the school grade becomes higher.

4. *Simple, Partial and Multiple Correlations of Achievement with Stanford Mental Age, Group Tests and School Attitude.*—The variable "school attitude" was obtained by averaging the judgments of from six to eleven members of the school staff. Using a rating scale of five steps, the teachers independently rated the pupils they knew well for a composite of traits such as *application*, *earnestness*, *willingness*, *effort.* The variable "group test" is the average of the correlations of all tests except the Myers. The criterion is the composite of achievement in school subjects. The r's given are the averages of the correlations for grades IV, V and VI. These, together with the partial and multiple correlations (Kelley's formulæ), are given in Table VII.

TABLE VII.—SHOWING THE SIMPLE, PARTIAL AND MULTIPLE CORRELATIONS BETWEEN VARIABLES (1) COMPOSITE OF ACHIEVEMENT (2) STANFORD MENTAL AGE (3) MEAN GROUP TESTS, AND (4) SCHOOL ATTITUDE

1 Simple correlations	2 Partial correlations first order	3 Partial correlations second order
$r_{12} = 0.54$	$r_{12\cdot 3} = 0.36$	$r_{12\cdot 34} = 0.32$
$r_{13} = 0.52$	$r_{12\cdot 4} = 0.47$	$r_{13\cdot 24} = 0.31$
$r_{14} = 0.32$	$r_{13\cdot 2} = 0.32$	$r_{14\cdot 23} = 0.12$
$r_{23} = 0.55$	$r_{13\cdot 4} = 0.47$	
$r_{24} = 0.40$	$r_{14\cdot 2} = 0.14$	
$r_{34} = 0.30$	$r_{14\cdot 3} = 0.21$	4
	$r_{23\cdot 4} = 0.49$	Multiple correlations
	$r_{23\cdot 4} = 0.49$	$r_{1\cdot 23} = 0.605$
	$r_{24\cdot 3} = 0.30$	$r_{1\cdot 234} = 0.611$
	$r_{34\cdot 2} = 0.10$	

From column 1 it appears that, in these grades, the Stanford Mental Age gives about the same simple correlation (0.54) with achievement as the average Group Test (0.52). School Attitude gives a considerably lower coefficient—0.32.

What we want to know more exactly is whether these three variables correlate with achievement by measuring very much the same group of abilities, or whether each contributes something unique, so that by properly combining them, the composite will give a correlation much higher than that given by any one alone.

First, the partial correlations of the Stanford Mental Age and the Group Tests with achievement will be considered. The correlation of Achievement with Mental Age is 0.54, and the partial correlation (See column 2, Table VII) between Achievement and Mental Age, with Group Tests eliminated, is 0.36, and the partial correlation of Achievement with Mental Age, both Group Tests and School Attitude held constant is 0.32. This means that, to a considerble extent, the Mental Age and Group Tests measure the same group of abilities, although each measure certain abiiities, correlated with achievement, which are unique. The practical value of each is perhaps more clearly indicated by the Multiple Correlations (column 4, Table VII). Here it appears that when the Stanford Mental Age and Group Tests are perfectly combined by use of weights obtained by the regression equation, the multiple correlation is 0.605 as compared to 0.54 or 0.52, which Mental Age or the Group Test, respectively, alone gives. The addition of 0.07 or 0.09 to correlations of these magnitudes is important.

The variable "School Attitude" has received so much discussion in recent literature that it merits careful consideration.

School Attitude is positively associated with Mental Age ($r = 0.40$) and with Group Tests, although not quite so closely ($r = 0.30$). The correlation of School Attitude with Achievement (0.32) is little higher than that between School Attitude and Group Tests. It is possible that this correlation with achievement is wholly, or almost wholly, due to the fact that "School Attitude" as judged by our teachers, and intelligence as measured by our tests are identical, in part. A careful study of the various partial correlations shows this to be true. When the elements of Mental Age and Group Tests which are identical with School Attitude are eliminated ($r14 - 23$), the residual of school attitudes gives a correlation with Achievement of but 0.12. The unique factors add very little to a composite of Mental Age and

Group Tests when each is properly weighted: the multiple r, achievement with (Mental Age + Group) is 0.605, and the multiple r, achievement with (Mental Age + Group × School Attitude) is 0.611.

It should not be considered that these facts greatly minimize the importance of school attitudes. The significant thing is that in so far as the School attitudes affect achievement in school work, they are almost completely measured by the intelligence tests. The Stanford-Binet measures these attitudes a little better than the group tests (the partial correlations $r_{14.2} = 0.14$, $r_{14.3} = 0.21$); both tests together account for them almost entirely.

These facts may be taken by some to support the doctrine that *interest*, *application* and the like, are in general, a sympton of intelligence—a *result* of successful functioning rather than a *cause* of it. The truth may be, however, that intelligence, as many seem to conceive it, and application, etc., as here estimated, are quite distinct, and that each contributes to success in the tests just as it does to success in school work.

The perfectly combined results of group tests, individual tests and school attitude fall far short of perfect correlation with achievement. What factors are reponsible cannot be discerned, at present, with certainty. The restriction of range in our groups and various defects in the instruments tend to reduce the correlations.

Two other possibilities will be suggested in the next section: (1) Specialization of abilities in school subjects, and (2) differences in the emphasis placed upon subjects by different schools.

(To be concluded in May)

ONE ELEMENT IN THE PROBABLE ERROR OF A MENTAL AGE MEASUREMENT

MARGARET V. COBB

Institute of Educational Research, Teachers College, Columbia University

The point I wish to make in the following paragraphs is this: In any measurement of mental age made by means of a scale whose units are relatively large (such as the Binet Scale, in which the smallest unit is 2 months) there is, in addition to other sources of error, an error due to the fact that the scale proceeds by steps instead of being continuous. The size of this error is dependent on the size of the steps in the scale, its least possible median value being that of half of one step. In the Stanford Revision of the Binet Scale, in which the unit varies from 2 months at the lower end to 6 months at the upper end of the scale, the minimum for this median error (or probable error) of a single measurement varies accordingly from 1 month at the lower end to 3 months at the upper end of the scale. Moreover, the error from this source always lowers the obtained mental age from its true value.

To make this clear, we may suppose a case in which this element of error is as small as possible—the case in which mental growth regularly parallels the scale, so that once every 2 months a new test is passed. Take for instance a normal child of 6 who is able to pass all the tests through Year VI, and none beyond this. His true IQ we will say, is 100 precisely. Just at the end of each 2-month period he becomes able to pass one more test (in Year VII or beyond) than he could before. Then at 6 years 2 months he would (if tested) measure 6–2, at 6 years 4 months he would measure 6–4, at 6 years 6 months he would measure 6–6, etc., and at each of these points his IQ would be 100. Between these points also his true IQ is 100, but would our measurement give this result? Suppose he were tested at the age of 6 years 1 month. He has not yet become able to pass another test—he is still a month away from it. He would still measure 6–0, and his apparent IQ would be 99. Just before he becomes 6–2, he would still test 6–0, and his apparent IQ would be 97. In other words, since there is no possible mental age between, for instance, 6–0 and 6–2,[1] there must occur somewhere (just as the child becomes able to pass another test) two successive days on one of which his mental age is 6–0 and on the other of which it is 6–2. By our assumption, his chronological

[1] Half credit on test 4, Year VII (tying bowknot) permits a single exception to this statement.

TABLE I

After	CA	MA	IQ
2 weeks	6–0	6–0	100
4 weeks	6–1	6–0	99
6 weeks	6–1	6–0	99
8 weeks	6–2	6–0	97
2 months	6–2	6–2	100

age on these two days is 6 years 2 months. His apparent intelligence quotient, which at 6 years exactly was 100, has gone down week by week to 97. It now jumps to 100 again. (See Fig. 1.)

When measured by a scale whose unit is 2 months, a child's mental development must appear to proceed discontinuously, shooting up suddenly at least 2 months at a time. When it happens that he reaches the point of success in two or three tests simultaneously, his gain will appear as 4 months, or as 6 months, at one time. It is not probable that this often happens in a single day, but perhaps not infrequently within a week or two, so that mental age, as measured, may legitimately increase 6 months or more within a few weeks,—or, on the other hand, 6 months or more may go by in which it happens that a child passes by none of these critical points which increase his measured mental age. In the latter case, for six months or more his IQ appears to be decreasing, for his chronological age is increasing, while his measured mental age is constant.

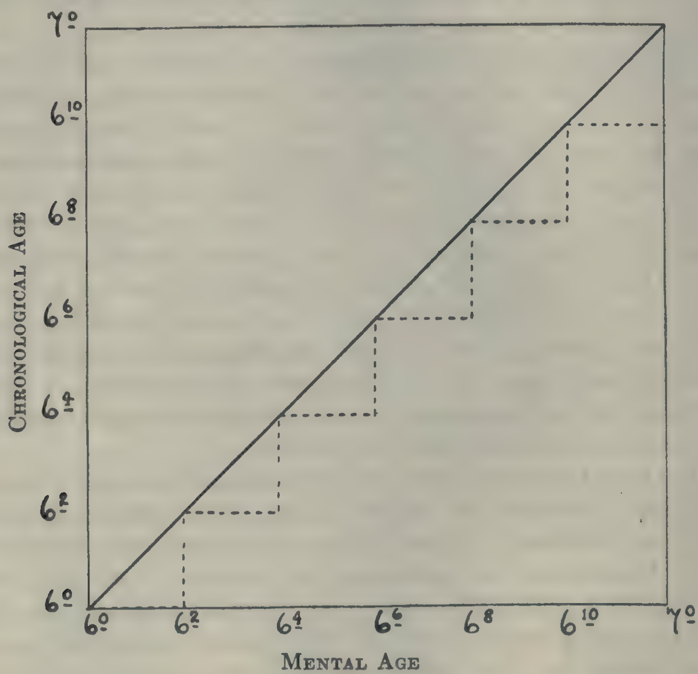
We are not to assume, however, from this discontinuity of measured mental age, dependent on the scale, that mental growth in children is itself discontinuous. What evidence we have points to continuous growth, though the rate may not be constant. Our child whose mentality becomes 6–2 just when he is 6 years 2 months old, is mentally only 1 day away from 6–2 the day before, even though his measured mental age is 6–0. The measurement on that day may be said to have an "error" of 2 months—an error whose existence may be laid to the presence of steps in the scale, and whose size corresponds to their size. On the next day the corresponding error is 0, since he now passes the test. One month hence the error will be 1 month; 2 months hence, again 2 months; 2 months and a day hence it will again be 0. In this case, where a rate of growth has been assumed for the child that will correspond as closely as possible to the scale, so as to give the *least possible* error, there still remains an inevitable error which varies in

FIGURE 1.

———Relation between chronological age and true mental age, when IQ = 100 and growth is regular.

------Relation between chronological age and measured mental age, when IQ = 100 and growth is regular.

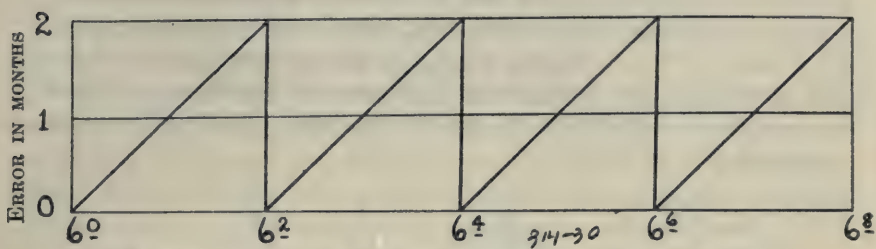
At any point vertical distance between two lines represents error due to size of steps in the scale.



regular periods from 0 to 2 months, as illustrated in the diagram below (Fig. 2), and always reduces the measurement.

It is obvious from this diagram that, since a child is equally likely to be examined at any age point, in half the cases the error will reach above and in half the cases it will fall short of the horizontal line drawn at 1 month—in other words, the median error here represented is 1 month. This is the minimum probable error due to size of scale units from Year III through Year X. Beyond Year X this minimum error increases as the steps of the scale increase in size, approaching its highest value of 3 months (due to this cause alone) when an individual who is growing is failing on those tests alone which appear in Year XVIII and count for 6 months each.

FIGURE 2.



CHRONOLOGICAL AGE AND TRUE MENTAL AGE (TRUE IQ = 100)

It must very seldom be the case, however, that our assumption is true, *i.e.*, that a child grows in just the way that will bring it about that he is able to pass a new test (*i.e.*, one on which up to that time he would have failed) at regular intervals. He may reach the critical point in regard to several tests within a few weeks, or within 6 months, and then go on for 6 months without passing any such points. The farther the departure from regularity in this respect, the greater becomes the range of error (and the median error) dependent on size of scale units. Until we know what degree of irregularity is usual, it is not possible to say exactly how great is the median error (or probable error) due to this cause; but until we have a scale with smaller units it can never at the lower levels be less than 1 month, nor at the upper levels less than 3 months.[1] It is not improbable that the true figures may be more than twice as large.

[1] When only half the scale is considered, the part of the PE under discussion is doubled. When the four starred tests instead of all six are used in each year, it is increased 50 per cent.

Here, then, is a factor which, together with change in type of content from one part of the scale to another, unavoidable variations in the giving of the tests, and personal reaction between examiner and child, is always present in the total PE of a mental age measurement. This total PE has been reported to range from approximately 3 months at the lower end of the scale to approximately 6 months at the upper end.

If, on the other hand, we wish to determine how great a difference might in extreme cases be brought about by this cause, an estimate only is possible. It is readily conceivable that occasionally a child may be tested by the standard procedure and obtain a mental age which, had he been measured a week or two later, would have been as much as a year higher. This, at chronological age 6, means 16 points difference in the IQ. Occurring in the first or in the second of two separated examinations, it could bring about a discrepancy, upward or downward, of 16 points if not more, and may account for the whole or part of some of the rather large "changes in intelligence quotient," the occasional occurrence of which has been reported by various investigators.

DISPARITY BETWEEN INTELLIGENCE AND SCHOLARSHIP

CHARLES LEONARD STONE

Dartmouth College

Even were intelligence and potential scholarship to correlate perfectly, there would always be cases of disparity between intelligence and scholarship grades: for the idlers, the men with excessive extra-curricular burdens, the men with unhealthy bodies or unhealthy philosophies of life skew the correlation on the one hand; and the men with unusual pertinacity or disproportionate absorption in scholarly performance upset calculations on the other hand. Any statement regarding the validity of a test as a predictive measure is premature, or speculative, until the disparity has been measured and some of the factors related to the disparity ascertained.

In a study of this problem with the class of 1924 at Dartmouth, the scholarship averages of the first semester were converted into percentiles, and likewise the intelligence ratings as determined by Alpha. The difference, in the case of each student, between the two percentiles was taken as the measure of discrepancy. On the basis of such computations, including 600 cases, it was found that over half of the class had discrepancies of less than 20 per cent, or 20 points on the percentile rating.

Discrepancy	Cases	Per Cent
0	13	2.2
1– 9	161	26.8
10–19	141	23.5
20–29	96	16.0
30–39	67	11.2
40–49	60	10.0
50–59	36	6.0
60–69	15	2.5
70–79	7	1.2
80–89	4	0.7

Of the 62 cases with discrepancies of 50 per cent or over, 53 remained in college during the second semester. It was decided to give tests to these men to measure their assimilation of the environment, their tension and persistence of work, and their power of aggressiveness. As a control group, 45 men who showed not over 2 per cent dis-

crepancy between Alpha score and scholarship were summoned with the other group. Of the 98 men summoned, 87 appeared to take the tests.

To measure tension of their normal work, a modification of the Downey will-profile test was given: The subject was asked to write the phrase "The American Legion" at his usual rate and in his usual style, then to write it as rapidly as possible. The speeded time divided by the normal time gave a quotient which was regarded as a measurement of his usual tension. To measure perseverance, the same phrase was written as slowly as possible, and this time divided by the normal time.

For the measurement of the degree to which he had assimilated the environment, an abbreviated form of the Rosanoff, Martin, and Rosanoff High Standard Frequency Test was given. As each of the 50 words was read, the subject was asked to respond with the first word that came to his mind. This response was evaluated on the basis of the number of men out of a group of 100 eminent men of science who gave this identical response. The total score is merely the sum of these values.

For the aggression measurement the following procedure was used. Three tests were given in this order, one to measure fear, one to measure, sex, and one to measure pugnacity or aggression. With three such tests, the first would surely render all necessary service as a shock-absorber or technique-developer, so that the later tests would be a more accurate measure of the instinct. The tests themselves consisted of stories of about 290 words meant to appeal conspicuously to the instincts, respectively, of fear, sex, and aggression. Scattered variously throughout the story were 20 words obviously unrelated to the story. The subject was told to cross out every word not belonging to the story. The time taken on each test was assumed to be a measurement of the strength of the instinct: for the more appealing the story the more likely were excursions into related imagery; and the greater disconcertion caused by the intruding words would, in all probability, make reorientation to the trend of thought more difficult and slow. The time taken by the subject on the aggression test was then divided by the total time of the three tests combined. This quotient seemed to be the most accurate measure: For the normal rate of reading would vary in the different individuals, and such variation between individuals would be eliminated by a quotient. If the introspections of the subjects may be trusted, these tests did not cause

the desired effects: Many reported that they were unaroused by the story as they rushed through it to delete the irrelevant words, but were thrilled after the test as they reread the passage for the story.

In the correlations there were marked divergences between the control group and the discrepancy group in only two of the five measurements:

	Control	Discrepancy
With scholarship		
r normal speed of writing	0.097 ± 0.104	0.609 ± 0.064
r Rosanoff High Speed Frequency	0.632 ± 0.063	0.002 ± 0.103
With intelligence		
r normal speed of writing	0.123 ± 0.103	0.036 ± 0.101
r Rosanoff High Speed Frequency	0.643 ± 0.062	0.181 ± 0.100

Since the group with discrepancies included both those with scholarship averages superior and those with scholarship averages inferior to Alpha rating, scores in each of the measurements were correlated with the differences between percentiles of scholarship and intelligence, from negative to positive. The data showed that normal speed of writing correlated 0.131 ± 0.100 with scholarship superiority; tension $0.185 + 0.98$ with scholarship superiority; perseverance 0.055 ± 0.101 with scholarship superiority; high standard frequency 0.131 ± 0.101 with intelligence superiority; and aggressiveness 0.016 ± 149 with scholarship superiority.

All of these latter correlations are low, but their tendencies seem to justify the assumption that discrepancies in favor of scholarship may be measured by some sort of perseverance or tension test, and discrepancies in favor of intelligence by some test which measures facility in absorbing the environment. Such tests deserve investigation.

No attempt was made to gather data regarding men's estimates of their industriousness or facility. Such data might be of accessory service, but should of course be subject to the usual scientific prescription, *cum grano salis*.

Possibly a "shock-absorber" for the intelligence test might be constructed on the basis of the Downey tension idea, with certain material to be copied at normal speed and other material to be copied

at full speed, the score being measured in units of amount copied rather than in units of time. Likewise much material from the Rosanoff High Standard Frequency Test might be incorporated into the information test. In all, it seems highly possible to incorporate within a single group test material to measure both mental abilities and the chief factors accessory to those abilities.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

Intelligence Tests

Intelligence Tests and the Classification of Pupils. F. S. Breed and E. R Breslich. The School Review, 1922, March, 210–216. Report of an investigation conducted to determine how well intelligence tests predict the educational achievement of high school pupils. Tests used were Otis, Chicago, and Terman Group Tests; Hotz Algebra Tests; and special school tests in arithmetical ability. Tests provide basis for temporary classification only.

Intelligence as a Factor in School Progress. I. N. Madsen. School and Society, 1922, March 11, 283–288. Retardation and acceleration among 12,182 children, grades III to XII as shown by Army Alpha and Haggerty Group Tests.

A Comparison of the Brightness of Country and City High School Children. James H. Hinds. Journal of Educational Research, 1922, February, 120–124. Results of administering the Otis Test to 454 city children and 59 small town children and 68 country children. Country child lower in mentality.

Prevention of the Lockstep in Schools. L. W. Cole. School and Society, 1922, February 25, 211–217. Differences in mental age among kindergarten and first grade children as shown by the Cole Vincent Group Intelligence Test for School Entrance. Comparisons with Binet Mental Age. Eight tables.

Suggested Studies in the Field of Mental Testing. Arthur S. Otis. Journal of Educational Method, 1922, February, 220–232. Enumeration and discussion of 8 research problems as suggestions to teachers and administrators. Help in undertaking the researches is offered by the writer.

Miscellaneous

A Method of Commensurating Mental Measurements. Harry S. Will. Journal of Educational Research, 1922, February, 139–153. A description of a new method called the "kental," for precise comparison of scores in a series of tests in various subjects. A comparison of the "kental" and the Galton percentile.

Administrative Problems Connected with Gifted Children. John C. and J. L. Almack. Educational Administration and Supervision, 1922, March, 129–136. Description of the selection of superior children in Eugene, Oregon. Problems met by administrators.

The New Knowledge of Spelling. C. H. Ward. The English Journal, 1922, February, 78–88. "Trouble Spots" in spelling in the high school. Causes and suggestions for improvement.

Vocational Guidance in the Junior High School. Frederick Schultz. Educational Review, 1922, March, 238–246. Suggestions for solving the problems of real vocational guidance.

Arithmetic Ability of Men in the Army and of Children in the Public Schools. Arthur Kolstad. Journal of Educational Research, 1922, February, 97–111. Comparison of the scores made on Test 2 of the Army Alpha by 2500 adult men, 632 children, grades 4 to 8, and 725 normal school students. Details shown in 5 Tables and 3 Diagrams.

Coercion and Learning. Wm. H. Kilpatrick. Journal of Educational Method, 1922, February, 233–239. Final installment of a popular discussion of the laws of learning.

Lost in Concrete Instances—Many Learners. Garry C. Myers, Journal of Educational Research, 1922, February, 135–138. A warning to teachers to make sure that pupils grasp the principles involved in concrete illustrations of abstractions.

Standards for the English Teacher. Allan Abbott. English Journal, 1922, February, 69–77. An address given before the National Council of Teachers of English. A proposal for setting up standard tests of teacher attainment. Details of program for preparing a series of standard tests.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

1. *Developing Mental Power*, a monograph of the Riverside series by G. M. Stratton,[1] is as delightful and stimulating a book for teacher as has appeared since James' Talks. Reviewing the extreme views on mental training, the mental disciplinarians and the doctrine of "contents," the author submits that both have missed the truth. It is urged that the mind is unified and although its parts are distinguishable they are closely related. The mind while particularized in ability is whole and fluid. The development of mental power is to be achieved by practice in control of the instincts, emotions, and the will, together with the acquisition of appropriate bodily and mental habits. Children must learn, through practice, to do the irksome, to be persistent and steady, to use caution and forethought, to develop reverence, taste, cheerfulness, courtesy, honesty. Many "exercises" for the care and training of the instincts, emotions and will are given, not on the disciplinarian assumption of the virtue of unpleasant and useless tasks since "the mind can strengthen on what is of service and delight, of which there is enough, without incessant treadmill work. Better to paint the ship, for discipline, than to knock rust off the anchor." Every teacher should spend an evening with this sane and stimulating essay which does not emphasize knowledge less but character more.

A. I. G.

---

2. *An Advanced Text in Systematic Psychology*.—Psychologists "are giving their time and attention almost entirely to the development of the experimental method and the discovery of individual facts by its means, to the serious neglect of the broader significance of these facts." In making this statement Mr. Moore, in the introduction of his book,[2] is stating an opinion frequently voiced by psychologists

---

[1] Stratton, H. M.: "Developing Mental Power." Houghton, Mifflin Co., Boston, 1922, pp. X + 77.

[2] Moore, Jared Sparks: "The Foundations of Psychology." Princeton University Press, Princeton, 1921, pp. XIX + 239.

themselves in recent conventions. Moore's work, aside from Boris Sidis' *The Foundations of Normal and Abnormal Psychology*, is about the only book attempting to evaluate the relations of psychology as a system, to the great problems of philosophy and of the natural and mental science. It differs markedly from the available histories of pscyhology and although written by a philosopher, it is primarily a discussion of the definitions and concepts of different schools of psychology and the postulates necessary for the construction of a scientific pscyhology, with special attention to the problems of parallelism, psychical causation and the subconscious. It is a good basal text for courses in systematic psychology.

A. I. G.

---

3. *A New Advanced Text on the Nervous System* by Tilney and Riley[1] is without doubt more comprehensive than any other printed in English. While it is written primarily as an introduction to the study of nervous diseases, it supplies many needs of the advanced student of psychology. It differs from the conventional texts in giving relatively large space to a discussion of the functions of nervous mechanisms. The various "syndromes" *i.e.*, complexes of symptoms, due to disease or destruction of nervous structures are described in detail together with treatments of the evolutionary significance, the relations, surface appearance, anatomy, histology, extirpation and stimulation experimental data with reference to each division of the system. The illustrations, photographic and diagrammatic, are numerous and admirably instructive.

The authors indulge infrequently in speculations concerning the nature, origin and localization of mental traits and such hypotheses offered are reached "by the direct and practical approach of clinical experience." "All areas of the brain outside of the frontal lobe are purely cognitive in their activity." As cognitude functions they discuss sensation, perception, "knowledge of external things." But every cognitive process "has attached to it some affective value." It is believed that the thalamus is the "site of the primary emotions and acts as that part of the brain primordially concerned in feeling tone." In the elaboration of feeling and primary emotions into "binary and tertiary combinations such as loathing, contempt, etc.,"

[1] Tilney, Frederick and Riley, Henry A.: "The Form and Functions of the Central Nervous System." Paul B. Hoeber, New York, 1921, pp. XXIV + 1020.

the procedure of McDougall in obviously followed. "*Knowing* must be activated by feeling before volitional reaction occurs." "If search were made among the various areas of the brain to determine the most likely place for the elaboration of these complex syntheses, the frontal lobe would doubtless be selected as best fitted for such process"—how reminiscent of Wundt's remarks nearly a half century ago!

Whether the reader will approve of Tilney and Riley's analysis of mental states or not, the book is a masterly treatise which may be studied with profit by the advanced student of physiological psychology.

A. I. G.

---

4. *Psychology in Industry.*—James Drever whose *Psychology of Everyday Life* has been recently reviewed in this journal is the author of a popular account of industrial psychology.[1] Chapters on intelligence testing, trade testing, fatigue, work and rest periods, motion study, external conditions, and the like, are written with an effort at clearness and simplicity. While the book contains no original contributions, it will serve as a useful introduction to study in this field.

A. I. G.

---

5. *A Readable Text for the General Worker in Mental Measurement.*[2] In this volume the author has brought together in unusually readable form practically all the techniques needed by the average worker in the field of mental measurement. The book is divided into three parts, each complete in itself, and followed by a list of supplementary readings. Part 1 (Chapters I–VI) deals with the place of scientific measurement in education and its uses in classification, diagnosis, teaching, evaluating efficiency of instruction, and vocational guidance. Part 2 (Chapters VII–XI) presents detailed instructions for the construction and standardization of tests. Part 3 (Chapters XII–XVII) is concerned with the application of statistical methods to test data and with the use of tabular and graphic methods in the effective presentation of results. An appendix gives the chief centers for the distribution of tests.

---

[1] Drever, James.: "The Psychology of Industry." E. P. Dutton & Co., New York, 1921, pp. XI + 148.

[2] McCall, Wm. A.: "How to Measure in Education." The Macmillan Co., New York, 1922, pp. XII + 416.

Forty-eight Tables and 35 Diagrams interspersed throughout the text illustrate clearly the principles involved and the methods used.

CECILE M. COLLOTON,
The Lincoln School of Teachers College.

---

6. *A New Yearbook.*[1]—The National Society for the Study of Education has presented a timely discussion of intelligence tests in the Twenty-first Yearbook. Thorndike's introductory chapter on "Measurement in Education" is followed by Colvin's on "The Principles Underlying the Construction and Use of Intelligence Tests." In the first part of the next chapter, Rugg has presented elementary methods of treating test data so lucidly that "he who runs may read," and then demonstrates his versatility by devoting twelve pages to a critical discussion of the newer developments and applications of statistical methods, supplemented by a classified and annotated bibliography. In Chapter IV Whipple presents in tabular form condensed information concerning 46 tests, under the following headings: The compiler, composition, range of ages or grades covered, time to apply, number and nature of test elements, publisher, prices and references.

In Part II of the Yearbook the various practical uses of intelligence tests are discussed. The introductory chapter is general. Chapters II and III show how the intelligence testing program functions in the schools of Detroit and Jackson, Michigan respectively. The remaining chapters deal successively with the measurement of intelligence in the lower primary grades, the elementary school, the junior and senior high schools, normal schools, colleges and universities.

The Yearbook is an exceedingly valuable compendium and guide for those who realize the place of group tests of intelligence in the administration of schools. It should also stimulate critical selection and evaluation of tests and methods of treating data.

The Committee[2] is to be congratulated both on the quality and

[1] Intelligence Tests and Their Uses. *The Twenty-first Yearbook of the National Society for the Study of Education.* Prepared by the Society's Committee and edited by Guy Montrose Whipple. Public School Publishing Company, Bloomington, Illinois, 1922, pp. IX + 289.

[2] The Committee was composed of the following members, all of whom contributed to the Yearbook: Stephen S. Colvin, Chairman, Helen Davis, Bessie Lee Gambrill, Henry W. Holmes, Warren K. Layton, W. S. Miller, Rudolph Pintner, Agnes L. Rogers, Harold Rugg, M. R. Trabue, E. L. Thorndike, G. M. Whipple.

range of the materials it has brought together. The tone of the discussion is at no point controversial. The Yearbook should demonstrate the significance of the movement in question and save it from the disastrous effects of extravagant claims and fears.

L. Z.

---

7. *A Manual for Case Study* which is extremely compact yet definite and complete has been recently published by the California Bureau of Juvenile Research.[1] Comprehensive instructions, with sample data, are given for taking case histories of intelligence, temperament, physical condition, moral character, conduct, amusements, education, home conditions, etc., etc. Chapters are devoted to the scope and meaning of social case investigation, methods (such as interviews, correspondence, inspection of records, etc.), methods of evaluating data, the use of charts, symbols, etc. It contains tables of norms of height, weight, complete sample histories and excellent lists of references on each topic discussed. This booklet, which can be purchased for a quarter, merits use in normal school and college classes and could be studied profitably by educational workers in general.

A. I. G.

[1] Williams, J. Harold and Others: *Whittier Social Case History Manual. Whittier, Cal.* California, Bureau of Juvenile Research, *Bulletin* No. 10, Dec. 1921, pp. 98.

ADDITIONAL PUBLICATIONS RECEIVED

A. Publications in Educational and Applied Psychology

Book, William F. *The Intelligence of High School Seniors.* New York: The Macmillan Co., 1922, pp. XVIII + 371.

Brooks, Fowler Dell. *Changes in Mental Traits with Age: Determined by Annual Re-tests.* New York City: Teachers College, Columbia University, 1921, pp. 86.

Drever, James. *The Psychology of Everyday Life.* New York: E. P. Dutton & Co., pp. IX + 164.

Drever, James. *The Psychology of Industry.* New York: E. P. Dutton & Co., pp. XI + 148.

Dunn, Fannie Wyche. *Interest Factors in Primary Reading Material.* New York City: Teachers College, Columbia University, 1921, pp. 70.

Hingley, R. H. *Psycho-Analysis.* New York: Dodd, Mead & Co., 1922, pp. VI + 190.

MacPherson, William. *The Psychology of Persuasion.* New York: E. P. Dutton & Co., pp. 256.

Toops, Herbert Anderson. *Trade Tests in Education.* New York City: Teachers College, Columbia University, 1921, pp. VI + 118.

Voelker, Paul Frederick. *The Function of Ideals and Attitudes in Social Education.* New York City: Teachers College, Columbia University, 1921, pp. 126.

B. Mental and Educational Tests

Baker, Harry J. and Kaufmann, H. J. *Detroit Kindergarten Test.* Yonkers: World Book Co., 1922.

Breitweiser, J. V. *The Thorndike College Entrance Tests in the University of California.* Berkeley: University of California Press, 1921, pp. 17.

Otis, Arthur S. *Otis Self-administering Tests of Mental Ability.* Yonkers: World Book Co., 1922.

Stenquist, J. L. *Stenquist Mechanical Aptitude Tests.* Tests I and II. Yonkers: World Book Co., 1922.

C. Publications in the General Educational Field

Jordan, Arthur M. *Children's Interests in Reading.* New York: Teachers College, Columbia University, 1921, pp. 143.

Judd, Cornelius D. *The Summer School as an Agency for the Training of Teachers in the United States.* Nashville: George Peabody College for Teachers, 1921, pp. 88.

Kentucky Educational Commission. *Public Education in Kentucky.* New York: General Education Board, 1921, pp. IX + 213.

The Lincoln School of Teachers College. *The Students Councils.* New York City: The Lincoln School of Teachers College, 1922, pp. 36.

The Lincoln School of Teachers College. *Some Uses of School Assemblies.* New York City: The Lincoln School of Teachers College, 1922, pp. 70.

National Society for the Study of Education. *The Twenty-first Yearbook. Intelligence Tests and Their Use.* Bloomington: Public School Publishing Co., 1922, pp. IX + 288.

Phelps, Shelton. *The administration of County High Schools in the South.* Nashville: George Peabody College for Teachers, pp. 153.

Rosenberger, Noah Bryan. *The Place of the Elementary Calculus in the Senior High-School Mathematics.* New York City: Teachers College, Columbia University, 1921, pp. VII + 80.

Sharp, Lawrence Alexander. *The Present Status of Rural Teachers in the South.* Nashville: George Peabody College for Teachers, pp. 83.

Stone, Clarence R. *Silent and Oral Reading.* Boston: Houghton-Mfflin Co., 1922, pp. XVIII + 306.

Storm, Ashley V. *How the Land-grant Colleges are Preparing Special Teachers of Agriculture.* Nashville: George Peabody College for Teachers, pp. 136.

Thomas, Frank W. *Training for Effective Study.* Boston: Houghton-Mifflin Co., 1922, pp. XVIII + 251.

University of Pennsylvania. *Eighth Annual Schoolmen's Week Proceedings.* Philadelphia: University of Pennsylvania, 1921, pp. 325.

Webb, Hanor A. *General Science Instruction in the Grades.* Nashville: George Peabody College for Teachers, 1921, pp. 105.

William McGuffy School. *Studies in the Geography of Europe.* Miami: University of Miami, 1920, pp. 85.

WILSON. G. M. *Connersville Course of Study in Mathematics.* Baltimore: Warwick & York, Inc., 1922, pp. 110.

D. NEW SCHOOL TEXTBOOKS

ANDERSON, ROBERT F. *The Anderson Arithmetic.* Boston: Silver, Burdett & Co., 1921, Book I pp. IX + 274; Book II pp. IV + 282; Book III pp. VI + 312.

HILL, HOWARD COPELAND. *Community Life and Civic Problems.* Boston: Ginn & Co., 1922, pp. XX + 528 + XXXIII.

MCCARTHY, JOHN DALY. *Health and Efficiency.* New York: Henry Holt & Co., 1921, pp. VIII + 262.

STARCH, DANIEL AND MIRICK, GEORGE A. *The Test and Study Speller.* Boston: Silver, Burdett & Co., 1921, Book I pp. VIII + 89; Book II pp. XVI + 63; Book III pp. XVI + 64.

E. MISCELLANEOUS PUBLICATIONS

Carnegie Foundation for the Advancement of Teaching. *Sixteenth Annual Report of the President and of the Treasurer.* Boston: The Merrymount Press, 1921, pp. 205.

DUDLEY, W. H. *Organization for Visual Instruction.* Bureau of Education, Washington. Bulletin No. 7, 1921, pp. 24.

General Education Board: *Annual Report* 1920–1921. New York: General Education Board, pp. XI + 129.

HEALY, WILLIAM. *The Practical Value of Scientific Study of Juvenile Delinquents.* Washington: U. S. Dept. of Labor, Children's Bureau, 1921. Publication No. 96, pp. 31.

KEYSER, CASSIUS J. *Mathematical Philosophy.* New York: E. P. Dutton & Co., 1922, pp. XIV + 466.

MOEHLMAN, ARTHUR B., THOMAS, J. F. AND ANDERSON, H. W. *An Analysis of the* 1922–1923 *Budget Requests of the Board of Education, City of Detroit.* Detroit: Board of Education, pp. 63.

The Relation of the Federal Government to Education. Urbana: University of Illinois, 1921, pp. 110.

Women's Educational and Industrial Union, Department of Research.

Old Age Support of Women Teachers. Boston: Women's Educational and Industrial Union, 1921, pp. 120.

F. Publications of United States Bureau of Education

1. List of Publications available 1921.
   - Bulletin No. 5, 1921, *Part-time Education of Various Types.*
   - Bulletin No. 6, 1921, *Opportunities for Study at American Graduate Schools.*
   - Bulletin No. 9, 1921, *Present Status of Music Instruction in Colleges and High Schools.*
   - Bulletin No. 10, 1921, *The Visiting Teacher.*
   - Bulletin No. 11, 1921, *Pharmaceutical Education.*
   - Bulletin No. 13, 1921, *The Housing and Equipment of Kindergartens.*
   - Bulletin No. 14, 1921, *Education of the Deaf.*
   - Bulletin No. 15, 1921, *Medical Education,* 1918–20.
   - Bulletin No. 16, 1921, *Special Features in the Education of the Blind During the Biennium* 1918–20.
   - Bulletin No. 17, 1921, *Educational Boards and Foundations,* 1918–20.
   - Bulletin No. 19, 1921, *Kindergarten Education,* 1918–20.
   - Bulletin No. 20, 1921, *Developments in Nursing Education since* 1918.
   - Bulletin No. 21, 1921, *Higher Education,* 1918–20.
   - Bulletin No. 29, 1921, *Monthly Record of Current Educational Publications.*
   - Bulletin No. 31, 1921, *Monthly Record of Current Educational Publications.*
   - Bulletin No. 39, 1920, *Facilities for Foreign Students in American Universities.*
   - Bulletin No. 48, 1920, *Statistics of State Universities and State Colleges.*
   - Health Education Bulletin No. 10—*Suggestions for a Program of Health Teaching in the Elementary Schools.*

G. Miscellaneous Pamphlet Material

1. American Academy of Political and Social Science. The Annals, November, 1921. *Child Welfare.* 10 Depot Street, Concord, N. H.

2. Blumgart, Leonard. *Observations on Maladjusted Children.* National Committee for Mental Hygiene, New York City, 1921.
3. Goodlander, Mabel R. *Education Through Experience: A Four Year Experiment in The Ethical Culture School.* Bureau of Educational Experiments, 16 West 8th Street, New York City, 1921.
4. Martin, Lillien J. *Pedagogical Hints from the Results of a Survey of a San Francisco Public School for Delinquent Boys.* Martin Mental Hygiene Publication, No. 5.
5. *United States Interdepartmental Social Hygiene Board,* Annual Report for 1921.

H. O. R.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Vol. XIII May, 1922 No. 5

RESEARCH VERSUS PROPAGANDA IN VISUAL EDUCATION

FRANK N. FREEMAN

University of Chicago

An educational movement usually goes through three stages. The first stage in the adoption of new materials or methods is characterized by undiscriminating propaganda. This propaganda awakens a high degree of enthusiasm on account of its plausibility and the deficient criticism to which it is subjected. The enthusiasm with which the new movement is met leads to widespread adoption of the new devices.

The second stage of the movement is one of reaction and of decline. As the new device is subjected to the criticism which is derived from the experience of many teachers, it becomes evident that the claims which were first advanced for it were greater than could be justified. The reaction which ensues, as is usual in social movements, proceeds beyond the point of equilibrium in the opposite direction, and the movement falls into disfavor.

After a time the third stage sets in, due to the return of the pendulum toward the state of equilibrium. It is discovered that the truth lies between the over-enthusiasm of the first stage and the undue reaction of the second. The movement possesses some value for education but this value needs to be estimated by a careful study of its possibilities and its relationships to other educational processes.

The procedure by which the true value of new educational processes is usually determined is unnecessarily wasteful. The third stage might be reached in a much more direct fashion if the critique which is applied in the second and third stages were introduced at the beginning, and if the unsystematic trial of the method in the class room were supplemented, and in a measure superceded, by the more systematic and organized testing of scientific experimentation. By this means

progress could be made more steadily and without the wasteful process of large scale adoption of unproved methods.

We are in a particularly advantageous situation to pursue this more economical and scientific mode of examination because, in the first place, we are conscious of our educational history, and have before us many examples of the sort of reaction which has just been described. This consciousness should put us on our guard against the too rapid and uncritical adoption of new movements. On the other hand it should prepare us for the acceptance of progressive development and for the adoption of changes which have been thoroughly tested. The history of education indicates that education never stands still. The progress of invention and of social life outside the school demands that the school shall be adapted to these changes. History, then, both points out the necessity of the adoption of advanced procedure and warns against unsystematic and unscientific acceptance of every new proposal. In the second place, we have an advantage over previous generations in the possession of scientific technique of investigation. The rapid advance of laboratory experimentation and of statistical methods in the past generation gives us tools of research which have never before been available for testing out new movements in advance of their adoption in the school room.

The principles which have just been discussed apply with particular force to visual education. The various methods which are comprised under this head undoubtedly constitute an advance in educational procedure. They possess possibilities which should by all means be realized in the school room. On the other hand, there are signs that the advantages which visual education possesses are being somewhat over-estimated and viewed in an uncritical and unpsychological fashion. The uncritical enthusiasm which is being developed is expressing itself in the undiscriminating propaganda which is characteristic of the first stage of a new movement. How far this propaganda will lead to wide scale adoption before the method is sufficiently tried out, is at the present time uncertain. It is to be hoped that a careful critique will at least hasten the third stage of careful and discriminating estimate, so that the second stage of reaction may be omitted.

We may enforce the statement that undiscriminating propaganda is being made by a few examples. The most common statement made by advocates of visual education runs something like this: "It is estimated by psychologists that 90 per cent of our sensory experience

comes through the eye. It is also commonly accepted that the higher mental processes, such as memory, imagination and reasoning, are founded upon sensation. It follows, therefore that education should appeal chiefly to the sense of sight. Visual education promises to revolutionize educational procedure and to supplant the customary modes of presentation."

This argument is very plausible and is calculated to convince many people. It is, however, contrary to fundamental and accepted psychological principles. Psychologists do not concern themselves with estimates of the relative frequency of sensations of the different sense organs. I am reasonably familiar with a good many texts in psychology but I never met with a statement which is at all akin to the one which forms the foundation of the above mentioned argument. Sensation does not possess the immediate significance which is implied in this argument. There is, therefore, no point in trying to estimate the relative proportions which one type of sensation bears to the others.

A more particular examination will indicate some of the false assumptions which are contained in the argument. In the first place it is not true that an experience which is initiated by vision is wholly visual in character. In fact, the total experience may be very largely non-visual. The sensation which is the starting point of the experience may be a relatively minor part in the whole.

The emphasis upon the sensation leaves out of account the importance of its niterpretation. Sensation may mean vastly different things to different persons. We may illustrate this and other features of the analysis from the experience of witnessing a football game. Two persons watching a game may have the same sensory experience. The significance of this sensory experience, however, may be very little to one person and very great to another. The actions of the players may even seem ludicrous to a person who does not know their intent or purpose. The various signals which are used, the movements which are made in various stages in preparation of a play, possess a meaning only to a person who understands the game.

We must include also under interpretation certain features which would commonly be erroneously ascribed to sensation. While two persons might be exposed, in the photographic sense, to the same stimulus or set of stimuli, one person would see vastly more of what was going on than another. The novice, for example, would observe only a confused mass of players, while the trained observer would

notice which individual carried the ball, who formed the interference, what opposing player broke through and tackled the runner, and what tactics were followed by the rest of the players. All this would be the result of previous training and experience.

The sensation of the moment is supplemented, furthermore, by many other sensations which may easily be overlooked. In the example which we are considering the auditory sensations would form an important part of the total experience. The music of the band, the cheering and the signals by the players or the officials, as well as the undifferentiated sounds which emanate from the large crowd of people, form very important elements in one's consciousness. Without these, many of the actions would lose their significance, and much of the feeling or emotional tone would be absent. It is a question whether a person would not get a richer experience from the totality of the other sensations than he would from the visual sensations without the others. The foregoing description has left out of account perhaps the most important group of sensory experiences. These are the result of the active or motor responses which are made in any real situation. The spectator at an athletic contest exhibits these motor responses to a marked degree. Many other factors in the total experience might, of course, be described. One's interests and relationship to the teams themselves or to the institutions which are represented by the teams has a determining factor in one's total attitude.

The illustration has been carried far enough to indicate that the sensation which may be thought of as initiating the experience constitutes but a small fraction of the total experience. The total experience is made up of many other sensations and of attitudes, ideas, and feelings which are the product of much previous experience or training. The particular sense through which the present experience happens to originate may be of much less importance than it appears to be on the surface.

The relative unimportance of the sense stimulus is clearest in the case of intellectual processes. The same intellectual activity may be initiated by a variety of sense experiences. The comparative indifference of the initiating sensation may be summed up.

In the first place, a large portion of any experience is derived from other sensations than those of the chief sense which was stimulated. Many of these sensations may exist quite independently of the one which usually initiates them. In fact, imagination alone may serve to reproduce or to set up the greater part of the entire experience.

In the second place, the importance of the immediate sense is reduced by the fact that the greater part of the experience may be non-sensory in character. Intellectual activities, while they may be originally derived from simple sensory and motor processes, go far beyond their simple origins. A conclusive piece of evidence in support of this statement is the fact that certain individuals can carry on complete intellectual operations of a high order, who are entirely deprived of the senses which are usually considered the most important—sight and hearing. I refer, of course, to such persons as Helen Keller and Laura Bridgman. Even in these cases, some sensation is necessary as a starting point, but these cases demonstrate that the character of the sensation does not determine the character of the thought.

Finally, it is possible to translate from one sense to another. It has been found very difficult to determine whether or not it is more advantageous to learn by the use of one sense or another, because it is almost impossible to determine which sense is actually used by the individual learners. It is a psychological commonplace that the sense through which the presentation is made is not necessarily the one in which the person thinks. To take another example, every novel reader conjures up in his mind images of persons and scenes which are nearly as vivid and often more satisfactory than the pictures which might be presented to his visual sense.

The burden of the foregoing discussion is that it is a very hazardous procedure to argue regarding the character of the total experience from the character of the sensation which appears on the surface to be the chief element of experience. The particular sense through which experiences in general are initiated is not of paramount importance. It is, of course, true that certain senses may possess advantages in particular cases, and it is furthermore unquestioned that certain special experiences can only be initiated by particular sensations. For example, music is dependent upon hearing, and the appreciation of painting is dependent upon sight. These, however, are special cases, and no wholesale argument can be based upon them. It is necessary rather that each case be examined for itself, in order that it may be determined what the most advantageous type of sensory stimulus may be. The thesis of this paper is that the various problems of presentation must be treated as a series of special cases and that each must be decided on its merits.

The foregoing discussion furnishes a criticism of certain psychological arguments which have been presented in support of visual education.

That this criticism is justified is indicated by the results of the careful examination of visual methods in a recent experiment. This experiment has been carried on by F. D. McClusky, who will shortly report the detailed results himself. I may, however, anticipate his report by citing some of the very general conclusions which are to be drawn from it. The investigation included over 700 children. It consisted in the comparison of the results of different forms of presentation of lessons in history, geography, and natural science. In each case the comparisons were made with great care and with an observation of the various checks necessary to secure valid results. Comparison was made between different modes of visual presentation, such as motion pictures and slides, with combinations of oral and visual presentation and with oral presentation alone.

The results of this study indicate that there is no justification for the adoption of the visual methods in exchange for those which are at present in use, on the basis of any wholesale conception of the superiority of vision. In fact, if the examples which were studied are to be taken as the sole basis of an estimate, one would have to conclude that visual methods possess little, if any superiority, and that the newer motion picture methods have no advantage over the older visual methods. The chief reason for not accepting the results of study at their full value and adopting this conclusion, is that these newer methods probably possess potential values which have not yet been fully realized. In order to determine more exactly what these potential values are, we need still broader investigation. Present investigation, however, is entirely adequate to constitute a complete refutation of any sweeping claims for visual education on the ground of general supremacy of visual sensations.

It is obvious that the problems of visual education are not solved at the present time. We should not expect them to be solved if we adopted a rational attitude toward the matter. The limitations of the visual method, which appear as a result of these experiments, would appear sooner or later as a result of their general use in the class room. It is, therefore, in the interests of the progress of visual education that its limitations be pointed out early. Experimentation is desirable also to indicate the direction in which visual methods should be developed, in order that their greatest possibilities may be realized.

It is possible by a more careful psychological analysis to determine something of the special uses and advantages of visual education, and something of its limitations in advance of experimentation. This

analysis, of course, must be regarded as somewhat tentative, but it has the merit of considering the problems in specific fashion rather than in the wholesale fashion in which they are often viewed. One of the limitations of the method grows out of the fact that only a certain type of meaning can be conveyed by objects which are presented to the eye. I refer, of course, to concrete objects and not to visual symbols, such as the printed word. The meanings conveyed by concrete objects or their pictorial representation must be of a rather concrete, simple kind. Such representation is not suited to convey the more subtle abstract or general meanings. These meanings are conveyed by language.

We are now in a period in which language is viewed with suspicion and disfavor. One may convince himself, however, of the necessity of language by observing its use in connection with motion pictures. Motion pictures themselves give the raw material of the experience, but the significance of this material is furnished by the captions or the reading passages. Let one refrain from reading these captions and he will be convinced of the comparatively large share of the meaning which is conveyed by them. It is true that certain crude type of meaning can be conveyed visually, such, for example, as physicial combat. This is probably the reason that fighting is so common an occurrence in the ordinary motion picture production. Consider, as another example, the representation of humor in motion pictures. Here again we find that a certain type of humor can be conveyed visually. This is made familiar by the "slap-stick, custard pie" style of humor. Aside from this, however, the laugh is nearly always elicited not by the picture itself, but by the caption. The limitation of visual presentation, in the type of meaning which it can convey must be kept in mind in estimating the usefulness of visual presentation for education.

A second limitation is that visual presentation in general, and especially motion pictures, dispenses largely with the personal influence of the teacher and with the social inter-action of the members of the group. This is noteworthy at a time when it is thought desirable to extend rather than reduce the teacher's influence in supervising the pupil's learning processes. This is an aspect of the matter which should be carefully considered. The teacher before the class can hold the attention of the pupils by eye, voice, and personal presence, and can determine, by watching the children, whether they are following the discussion, and thus adapt the pace to their own progress. It is a

common assertion that motion pictures hold the attention of pupils more strongly than do other forms of class exercise. McClusky's study indicates that this is to be seriously questioned. This is a matter which needs further experimentation, and on which it is necessary to be cautious in accepting the conclusions from the entertainment movies.

Over against these limitations are to be set certain probable advantages. It is undoubtedly true that pictures and visual stimuli generally possess a certain immediate appeal. This is an appeal which visual material shares with other sensory stimuli. It is an advantage of visual stimuli particularly because of the large amount of material which is susceptible to this mode of presentation. If we could present the same material directly to the other senses, we might find the sensory appeal to be as strong as in the case of vision. When we estimate this appeal of visual material, however, we usually compare it with presentation through language, either in print or oral speech. While it is true, as has already been argued, that presentation through language is essential to give meanings of the more general or abstract sort, it is also true that for most persons, such presentation through language has somewhat less direct appeal than have sensory experiences. For the presentation of meanings which can be conveyed through sensory channels, therefore, it is desirable that concrete materials be employed.

In the next place, certain types of relationships may be most clearly apprehended when they are represented visually. Visual devices are particularly suitable for the representation of special relations. One may gain a much clearer notion of a geographic region from examination of a map than by any other means. The construction and operation of mechanical contrivances, furthermore, are better shown than described. It goes without saying that the graph is an unrivalled method of presenting certain types of relationships between facts, certain general comparisons and general trends and changes. Other types of changes may be represented peculiarly well by motion pictures. Noteworthy examples are the analysis of a rapid motion by a picture which is slowed down in the projection, and the representation of very slow movements by rendering them perceptible through speeding up the projection. All of these advantages are unquestioned and important. They indicate the direction which it would probably be most profitable for the development of visual methods to take.

Motion pictures, slides, models, and other visual materials share with text books the advantage of being the means of diffusing expert examples of presentation. H. G. Wells has emphasized this advantage in an exaggerated fashion in his famous articles on education. Pictorial representation is also of advantage in making widely available the working of rare or expensive apparatus. Similarly the performance of a difficult act by an expert may be analyzed and presented broadcast.

Experimental research should be devoted to the discovery of the types of educational material which are best adapted to visual presentation. An analysis, such as the foregoing, can simply point out the most probable lines of development on the basis of our general psychological insight. Such analysis needs, however, to be supplemented and verified by careful scientific procedure.

Experimentation should be applied also to the study of certain problems in the development of the visual method itself. The problems which are here mentioned relate particularly to motion pictures. One of these problems concerns the span of attention. The traditional material is organized so as to conform to the span of attention of pupils for whom it is intended. This is true of text book material and of oral lessons. This organization has been developed empirically through long periods of use in class room. It is possible to work it out more quickly and systematically by scientific experimentation.

Another problem which should be attacked is the best method of securing the attention of the pupils. The attractiveness of emotional films is sometimes exaggerated by pressing the analogy of films which are designed for entertainment. It is coming to be recognized, however, that educational films must rely upon different sources of interest. We cannot rely simply upon the sensory appeal which has already been mentioned. The primary problem is so to organize the film from the standpoint of intellectual apprehension that it may furnish, in addition to the sensory appeal, both intellectual stimulation and satisfaction. In other words, the presentation must be adapted to the intellectual capacities, interests, and activities of the pupils. A subordinate problem concerns the methods by which the attention of the class may be kept upon those features of the presentation which are central.

Related to this problem of attention is the question of the rapidity with which the units of thought or of subject matter are presented and the correlative question of amount of detail which should be included.

A given topic may be presented rapidly by stressing only the outstanding features, or it may be presented slowly with the addition of many details. It is sometimes mistakenly supposed that the omission of details simplifies the presentation. Finally, it is necessary to determine how much repetition and review is necessary in order to secure permanence of learning. All these questions are susceptible of experimental investigation.

In the interests of visual education, then, experimental investigation should be made to determine the type of educational subject matter to which it is best adapted, and the manner in which it may best be organized. Such a study will form the basis of steady and permanent progress. Unsound propaganda, on the other hand, will lead to more rapid initial progress, but this will be followed by a reaction which will result in slower progress in the end.

A FURTHER CRITERION FOR THE SELECTION OF MENTAL TEST ELEMENTS

J. CROSBY CHAPMAN

AND

A. BARBARA DALE

Department of Education, Yale University

In a short note[1] by one of the authors, attention was called to the necessity of a further criterion for the selection of mental test elements. Excluding considerations of inter-correlations, the two criteria by which mental test elements are most commonly judged are:

1. Increase of performance from age to age.
2. Coherence.

Neither of these criteria is very serviceable in investigating the extent to which performance in a test is conditioned by hereditary brightness or by mental changes produced largely by exposure to training influences. Obviously, the first criterion, increase of ability with age, tends to operate in the same direction as the training factor, while the coherence criterion is so crude as to offer no definite safeguard against this disturbing factor of environmental training. To quote from the note to which reference has already been made: "In thus establishing the validity of the test elements, we have been guilty of loading the dice in our own favor; we have made our task too easy. It is essential that we set up an additional criterion which will load the dice in the opposite direction. Whereas in the above situation, both factors, age and training, work in the same direction, we must set up a criterion in which they work in opposite directions."

The effect of environmental training as a complicating factor in intelligence testing is, of course, no new discovery; Binet, Chotzen, Stern, Terman, Pintner and Paterson, and many others have called attention to its influence. But no one has pressed the matter to its logical conclusion. What is required is a method of discrimination, other than by personal judgment, of the relative weight to be placed on the separate factors of intellect and training in the different elements of a test. It is generally agreed that intelligence must be measured in terms of the higher complex processes. It follows, therefore, that we must so arrange our experimental conditions that the higher processes

[1] Chapman, J. Crosby: An Additional Criterion for the Selection of the Elements of Mental Tests. *Journal Educational Psychology*, April, 1921, pp. 232–235.

can reveal their presence in the chronologically young but mentally bright child. The test elements which measure the presence or absence of these powers must be freed from non-essential experience factors. Otherwise we shall never get evidence of the presence of these processes in the young bright, not because they do not exist, but because, to show themselves, it is required that they be exercised with a facility not yet acquired by, or with material not yet imparted to, younger children. For example, the performance of the young, precocious child, even in a test such as the opposites, may be prejudiced, not necessarily by lack of speed of conceptual thought, but by inadequate vocabulary or possibly by the absence of facility in reading, due to lack of practice. The obvious fact that the bright child derives much more benefit than the dull child from precisely the same environment, or from much less environmental influence, does not by any means make the effect of environment and training negligible. The temptation is always present to estimate intelligence by mental tests so constructed as regards facility demanded and information required as to result in a measurement of factors which are exceedingly closely correlated with grade position. Whenever this is done, knowingly or unknowingly, we are, really, abandoning the mental test as the criterion and putting our trust in the high correlation which we know exists between grade and intelligence. On the surface we are relying on a short performance in the selected mental test elements but in reality we are putting our confidence in the long continued process of school selection to raise our correlation. This procedure, while unquestionably yielding, in an easy manner, fairly high correlations, does great injustice to individuals whose environmental opportunities are not essentially normal. Unfortunately, the present statistical procedure, correlation formulæ and mass treatment of data tend to burke the issue.

To return to the method of discriminating between native and environmental factors, let us suppose that a large number of children are given the same complete intelligence test, composed of various elements, some of which, typified by element A, call for a greater degree of native ability; while others, typified by element B, can be more satisfactorily performed by virtue of longer training and exposure to environmental stimulus. Suppose, moreover, that from amongst these children two groups, of differing chronological ages, can be so selected that each member of the one group is matched in total score by a corresponding member in the other group. Let us suppose that the first group, designated group O, consists of children over 13 years

of age, drawn from Grades VII and VIII while the younger group, group Y, is drawn from Grades III and IV, the age in every case being less than 10 years. With the data on the various tests from two such groups, it is possible to investigate the problem of chronological maturity and environmental influence: Group O will make up its total score more by good performance in Test B than in Test A; while the reverse will hold for Group Y. Hence in Test A the average score of the younger children will exceed that of the older; in Test B, compensatory marks will be gained by the older children. By these means it will be possible to rank a series of tests according to the superiority shown by the young bright pupils as compared with the old, dull pupils. This, then, will be their order of merit as tests for native intellectual endowment; the reverse order will rank the tests according to the degree in which they are dependent on the training factor. The fundamental assumption underlying this argument will be examined later in the paper.

The authors had at their disposal about 5000 National Intelligence Test Blanks (Series A) which had resulted from the administration of this test to Grades III to VIII in several elementary schools in Mount Vernon, N. Y. This material was furnished by W. H. Holmes, the Superintendent of Schools, to whom the authors wish to express their thanks. From this material were selected two groups which fulfilled the following conditions:

1. The parents of all children must be of American or British birth (*i.e.*, English speaking).

2. The children of the first group (Young Bright, Y. B.) must be less than ten years of age, drawn chiefly from Grades III and IV.

3. The children of the second group (Old Dull, O. D.) must be thirteen years of age or over and drawn from Grades VII and VIII.

4. The scores obtained by a member of either group must fall between 70 and 119. Otherwise, for the child of 9, the extreme low score does not represent brightness, nor, for the pupil of 13 years, does the high score represent dullness.

From these two groups a narrower selection was made by pairing, in the Old Dull group, each paper which could be matched in total score with a paper from the Young Bright group; if necessary a difference of one mark, and no more, was allowed between the totals of any one pair. In this way fifty pairs of papers, each pair evenly matched, were obtained. That is to say, we have two groups of identical content as regards total scores, the one composed of the lowest scoring

children of 13 years of age and over, and the other, of the highest scoring children of under 10 years. The methods of scoring and calculating totals were those dictated by the committee constructing the test. Interpreting the test strictly, the total scores being similar, this selection furnishes two groups of the same mental age, but of widely differing chronological age. Without strictly defining our terms we may, with reasonable accuracy, speak of an Old Dull group and of a Young Bright group, the average age of the first being 14 years 7 months, and the average age of the second, 9 years 3 months. There is, therefore, for each paired paper an average difference in chronological age of about 5 years, with a range from 3 to 8 years.

Each group was then separated on the basis of total scores into the intervals 70–79, 80–89, etc. The average score for the subjects within each of these intervals, for each test, and for totals, was determined for both groups. From these data, weighted according to the number of cases in each interval, averages were obtained for each group as a whole, in each of the tests. These results are presented in Table I, where Tests 1, 2, 3, 4, 5, are arithmetical reasoning, sentence completion, logical sequence, opposites and symbol-digit substitution test, respectively.

TABLE I.—SHOWING THE AVERAGE SCORES AND DISTRIBUTION OF THE TWO MATCHED GROUPS OF 50 CASES IN SUCCESSIVE INTERVALS OF 10 POINTS, THE FINAL AVERAGES AND RATIOS OF THESE IN EACH TEST

Interval	Frequency	Young Bright average scores						Old Dull average scores					
		Test I	Test II	Test III	Test IV	Test V	Total	Test I	Test II	Test III	Test IV	Test V	Total
70– 79	9	10.0	19.8	17.0	13.7	12.8	73.2	12.9	17.6	21.3	11.3	10.3	73.4
80– 89	9	9.8	19.6	20.1	19.4	15.9	84.8	12.1	22.3	23.3	9.4	17.3	84.6
90– 99	14	14.3	23.9	22.6	19.7	15.6	95.9	14.4	19.5	22.2	15.3	24.4	95.8
100–109	11	13.8	25.2	23.0	20.3	21.7	104.0	16.0	27.3	25.0	12.5	23.4	104.1
110–119	7	15.4	26.3	27.0	21.9	22.7	113.3	16.3	29.4	28.3	13.4	26.1	113.6
Final average..	50	12.7	23.2	21.8	19.0	17.5	94.1	14.3	22.8	23.7	12.6	20.6	94.1

	Test I	Test II	Test III	Test IV	Test V
Ratio of average scores $\frac{Y.B.}{O.D.}$....	0.87	1.02	0.92	1.51	0.85

At the bottom of the table the ratio of the performance of the Young Bright to the Old Dull is shown. Where the ratio is greater than unity there is a tendency for the test, as scored in the present procedure, to favor native intelligence; where the ratio is less than unity, the emphasis is rather on the training factor. It will be seen that Table I shows that in Test 4 (opposites), there is a very large ratio in favor of the Young Bright group. Test 2, the sentence completion, occupies an intermediate position, while Tests 3, 1, and 5 show decreasing requirements of brightness, and if interpreted strictly, favor training and chronological maturity. While there is a very clear discrimination between the opposites test on the one hand, and the substitution test on the other, it is only by a more elaborate examination of the data that the reliability of these results can be investigated.

To make this more complete investigation, the distribution in each test for each of the groups was made. For each of these distributions and for the totals, the average, median, and mean square deviation was determined. These are recorded in Table II and the data worked over in Table III.

TABLE II.—SHOWING THE CENTRAL TENDENCIES AND VARIABILITY OF SCORES AND AGES FOR THE TWO MATCHED GROUPS OF 50, TOGETHER WITH RELIABILITY DATA

Young Bright

Test	I	II	III	IV	V	Total	Age, Years, Mos.
Average	13.72	23.76	22.32	19.76	18.16	94.7	9.4
σ	3.55	4.39	4.58	4.01	5.48	13.4	0.6
σ	0.50	0.62	0.65	0.57	0.77	1.9	
Median	13.64	23.60	22.22	20.00	17.54	96.5	9.5
σ	0.62	0.77	0.81	0.71	0.96	2.4	
Old Dull							
Average	15.32	23.76	24.20	13.28	21.28	94.6	14.7
σ	3.45	6.48	6.00	6.29	8.82	13.5	1.1
σ	0.49	0.92	0.85	0.89	1.25	1.91	
Median	15.85	23.43	25.33	13.54	23.33	96.5	14.7
σ	0.61	1.15	1.06	1.11	1.56	2.4	

TABLE III.—SHOWING THE DIFFERENCES BETWEEN THE CENTRAL TENDENCIES IN THE TWO MATCHED GROUPS AND THE RELIABILITY DATA. ALSO RATIO OF AVERAGES AND MEDIANS FOR THE TWO GROUPS

Young and Bright Score *minus* the Old Dull Score in first four rows

Test	I	II	III	IV	V
Averages	−1.60	0.00	−1.88	6.48	−3.12
σ	0.70	1.11	1.07	1.06	1.46
Median	−2.21	0.17	−3.11	6.46	−5.79
σ	0.87	1.40	1.33	1.32	1.83
Ratio of averages $\frac{\text{Y.B.}}{\text{O.D.}}$	0.89	1.00	0.92	1.49	0.85
Ratio of medians $\frac{\text{Y.B.}}{\text{O.D.}}$	0.86	1.05	0.88	1.48	0.75

TABLE IV.—SHOWING THE CENTRAL TENDENCIES AND VARIABILITY OF THE 41 MATCHED PAIRS (ZERO SCORES ELIMINATED). ALSO A MEASURE OF RELIABILITY

Test	I	II	III	IV	V	Total	Age, Years, Mos.
Young Bright							
Average	14.17	24.44	21.98	20.17	18.59	97.50	9.5
σ	3.61	3.94	4.12	4.08	5.71	11.06	0.5
σ	0.56	0.61	0.64	0.63	0.89	5.46	
Median	14.43	24.15	22.71	20.55	18.00	98.25	9.6
σ	0.70	0.77	0.80	0.79	1.11	6.82	
Old Dull							
Average	15.54	24.05	23.76	13.9	23.55	97.38	14.8
σ	2.96	6.25	6.17	3.40	3.79	11.19	1.0
σ	0.46	0.98	0.96	0.53	0.59	5.52	
Median	15.90	23.54	24.83	13.64	24.17	98.25	14.7
σ	0.67	1.22	1.20	0.66	0.74	6.90	
Ratio averages $\frac{\text{Y.B.}}{\text{O.D.}}$	0.91	1.02	0.93	1.45	0.79		
Ratio median $\frac{\text{Y.B.}}{\text{O.D.}}$	0.91	1.03	0.91	1.51	0.74		

To meet the criticism that the results obtained for the median and the average scores of the two groups were affected by the presence of a few zero scores, this point was given further investigation. While no zero scores occurred in the younger group, it was discovered that there were four zero scores amongst the older group in Test 4, and six in Test 5.

To leave no doubt in the matter, those papers containing zero scores were eliminated in the older group, and the corresponding paired papers were taken from the younger group. This reduced the number of pairs to 41 and the process of calculating the central tendencies, etc. was repeated for these pairs. The results are combined in Table IV which is too similar to previous tables to need further comment.

As confirmatory evidence of the reliability of the above results, it may be well to add that, for three other pairs of 20 apiece, selected without eliminating the language question, the same general figures are obtained. The final averages for these pairs furnish the following ratios:

	Test I	Test II	Test III	Test IV	Test V
Ratio $\frac{\text{Y.B.}}{\text{O.D.}}$	0.92	1.04	0.97	1.44	0.78

Before any conclusions are drawn, one point should be made clear. Where the ratios expressing the performance of the Young Bright to the Old Dull differ greatly from unity it is safe to draw deductions. Where, however, the differences are small, nothing can be inferred. Suppose, for example, in a battery of four tests, one test was favorable to native intelligence, the other three being neither favorable to the young nor to the old group. In the first mentioned test the extra score of the young would have to be compensated by slightly lower scores on the three other tests, if the totals of the two groups were the same. We cannot, therefore, make any absolute quantitative estimates but must confine ourselves to stating, under the present system of scoring, the order of merit of the tests as measures of native intelligence, rather than of environmental training. These rank as follows:

(1) Opposites Test, (2) Sentence Completion, (3) Logical Selection, (4) Arithmetical Problems, (5) Symbol-digit Substitution.

It is now time to examine more closely the fundamental assumption upon which the procedure of this study is based. This may be stated as follows: A test element in which the performance of the Young

Bright exceeds that of the Old Dull is, except in unusual circumstances,[1] *ipso facto*, a superior test of native intelligence. In those tests where the Old Dull are superior to the Young Bright, this superiority can be explained in two ways. The first and most obvious explanation would be found in the fact that the Old Dull have been exposed for a longer period to training influences. The second explanation, which is less probable, would advance the hypothesis that the superiority of attainment of the Old Dull was caused not necessarily by longer exposure to environmental influence but, rather, by the maturation of innate mental powers due to greater chronological age. While it is impossible to show that the second explanation is erroneous, the following facts may be adduced as cumulative evidence against it. The two tests which exhibit the superiority of the Old Dull are the substitution and arithmetical reasoning tests. The former is somewhat dependent on acquired eye and hand co-ordination, while the latter is obviously much subject to school training. Furthermore, the test in which the young show their maximum superiority is the opposites, involving a somewhat high form of conceptual thinking, which is not subject to direct practice in the ordinary procedure of the school. It is also difficult to explain why the inner development due to the chronological age of the duller pupils should result in increased powers in one direction, with no similar increase in other directions. Certainly, the burden of proof rests with those who maintain the second position. The first explanation is much the simpler, fits in much better with the observed facts, and is supported by the theory of the general development of intellectual power.

In the light of these deductions, the acceptance of the total score as the criterion of equality of intelligence is subject to criticism. It may be said that these results are subject to the fallacy that our test of total intelligence is derived from the data, the validity of which we are examining. To meet this objection, it may be urged that as we are compelled to have some measure of intelligence, the total is probably the most reliable. Certainly as the test is usually employed the verdict depends on the total in all the tests.

If, as we are bound to assume when we employ a group test, the same total represents the same mental age, whatever may be the chronological age, this study shows that the mentality of the Young Bright is different from that found in the Old Dull. It also shows that

[1] For example where the Young Bright had recently practiced a function which the Old Dull had forgotten.

intellectual development is not marked by a simultaneous and uniform improvement in all types of mental function. The Old Dull subject of mental age x is the superior, equal, or inferior of the Young Bright of mental age x, mental age being measured by totals, according to the type of test used. This is surely a peculiar state of affairs.

It would seem therefore as though a very difficult problem must be faced by those who construct intelligence tests. If we select elements which are analogous to the opposites test, we shall thereby give advantage to a power which is found in the Young Bright, but to a less extent in the Old Dull of the same mental age, as measured by a group test. If, on the other hand, we use such tests as substitutions, we shall under-estimate the intelligence of the younger child and probably over-estimate that of the older. This gives rise to the anomalous situation that the score of a pupil, whether bright or dull, can be raised or lowered according to the elements which are selected for the test.

If the Young Bright and the Old Dull are scoring on different tests, this may have a significant effect upon the inter-correlations of tests; that is, upon the logic of partial correlation method, which partly determines the selection of tests. Tests which are free from environmental training influence may be eliminated because of high inter-correlations, in favor of tests showing lower inter-correlations produced by the environmental factor. If this is the case, upon what adequate psychological and sociological criterion shall selection be based?

A further point of debate raised by this study is the advisability, or even the possibility, of making comparisons of subjects of one chronological age with subjects of another chronological age. In the light of the results which have been obtained, is it fair to compare the performances of an 8-year-old with that of a 10- or 12-year-old? The use of mental age and the corresponding IQ coefficient assume the legitimacy of the procedure. In individual examinations, we give to precocious children of 9 years of age the tests which have only been devised and justified for children of, let us say, 12 years of age. For example, both criteria used by Terman only establish the fact that the test element is valid for children of approximately that chronological age in the neighborhood of its particular age position. Such procedure establishes no right to use this test, forthwith, on younger age groups that have not had the same period of environmental influence. Environmental influence is no theoretical factor which can be eliminated by shifting the tests from age group to age group, in the attempt to get a normal distribution of intelligence at each age.

It would appear that the use of the mental age method of measurement, while practically straight-forward, is subject to such inherent defects that for finer work in the realm of intelligence measurement it must eventually be displaced. A child of 8 and a child of 12 cannot be compared. It is an impossibility to select test elements which are not affected by the additional 4 years of environmental influence enjoyed by the latter child. Eventually we must state the performance of the x-year-old in terms of the performance of a large group of x-year-old children, using either percentiles, or distances in terms of sigma. Even with this precaution, the differences in environmental opportunity of the x-year-olds will make it difficult enough to select fair test elements. Under the present scheme such a selection is probably impossible to attain.

Summary

1. The question of the effect of environmental influence and training on performance in mental tests is again raised.

2. A new criterion is proposed which will rank tests with reference to the weight placed on hereditary brightness rather than on environmental training.

3. To apply this criterion, two groups, on the basis of totals in the National Intelligence Test, Series A, are selected, one consisting of the Young Bright and the other the Old Dull. The members of one group are paired, as far as totals in the test are concerned, with members in the other group.

4. It is shown that these two groups are identical in totals, score, in differing amounts, in the five tests constituting the examination.

5. The Opposites Test seems to depend, to a high degree, on native intelligence, while the Arithmetical Problems and Substitution depend more upon the environmental factor.

6. The significance of the above results and the effect on the selection of mental test elements is discussed.

7. The legitimacy of the present method of estimating intelligence by the IQ method is considered.

THE CORRELATIONS OF ACHIEVEMENT IN SCHOOL SUBJECTS WITH INTELLIGENCE TESTS AND OTHER VARIABLES (CONCLUDED)

ARTHUR I. GATES

Teachers College, Columbia University

5. *The Intercorrelations of School Subjects.*—Table VIII gives the mean intercorrelations of school subjects, and the correlations with MA and CA.

TABLE VIII.—INTERCORRELATIONS OF SCHOOL SUBJECTS. FIGURES ARE THE MEANS OF THE COEFFICIENTS FOR GRADES IV TO VIII INCLUSIVE

	Reading Comprehension	Reading Rate	Arithmetic	Spelling	Writing	Mental Age	Chronological Age
Reading Comprehension	......	0.85	0.24	0.45	0.06	0.49	−0.25
Reading Rate	0.85	......	0.22	0.38	0.21	0.46	−0.35
Arithmetic	0.24	0.22	......	0.24	0.09	0.30	−0.19
Spelling	0.45	0.38	0.24	......	0.02	0.31	−0.23
Writing	0.06	0.21	0.09	0.02	....	0.08	0.05
Mental Age	0.49	0.46	0.30	0.31	0.08		
Chronological Age	−0.25	−0.35	−0.19	−0.23	0.05		

Generally, the correlations of one subject with others are not high. Writing shows no association of significance with any variable here listed. Arithmetic correlates but slightly with other subjects. Spelling is more closely associated with reading than with other subject. Our criteria of General Achievement (a composite of Reading Comprehension, Reading Rate, Arithmetic and Spelling) are therefore seriously in need of study and evaluation. The measures of the particular subjects were extensive enough to warrant confidence in their reliability, although they may fall considerably short of perfect validity.

The two most likely explanations of the low intercorrelations among the school subjects are: (1) A specialization among the functions due primarily to original (inherited) aptitudes and (2) differences in the degree of possible achievement in each, depending largely upon the relative emphasis in teaching.

For both views there is evidence; a greater abundance for the former. The importance of the latter possibility is suggested by the studies of Hollingworth[1] and more recently by those of G. S. Gates,[2] which show an increase in the intercorrelation of even very narrow functions (tapping, color naming, etc.) as the subjects approach a practice limit. If we could secure a random selection of 12 year olds who had pushed achievement in reading, spelling and arithmetic to the very limit, the intercorrelations and the correlations with intelligence tests would probably be much higher than those now found. Whether arithmetic will yield high correlations with intelligence (assuming that we had a valid measure of general intelligence) in School X, Y or Z, probably depends considerably on what the school does in its teaching of that subject. Our data represent merely the facts for one school, but our impression is that it is rather unlikely that the correlations would rise near +1.00 even if each subject were developed to the limit. There is doubtless some specialization due to native endowment.

If this is true, two suggestions follow. The first is a question concerning the validity of the "accomplishment quotient," which is based on the assumption of no (or, at least, slight) specialization. The other suggestion is that we should attack the problems of discovering tests of native ability for each subject separately. We should have tests for native aptitude for arithmetic, writing, drawing, spelling, and so on.

6. *Group Tests and Stanford MA Correlations with Particular Subjects.*—In Part III, Section I, it was found that the more verbal the material, the higher the correlation with all subjects except arithmetic which was more closely associated with moderately verbal material. This is suggestive of a starting point in future research for tests of aptitude for the different subjects. The Stanford test and many group tests include materials which rank high, low, and at various levels on the verbal scale. If there is a specialization of native abilities, the result is a rather moderate correlation with a composite of all school subjects. Table IX gives a comparison of MA and Group Test correlations with the particular subjects, the figures representing the mean for Grades IV and VI.

---

[1] Correlations of Abilities as Affected by Practice. *Journal of Educational Psychology,* 1913, p. 405.

[2] Doctor's thesis (unpublished) in the Library of Columbia University.

TABLE IX.—CORRELATIONS OF THE MEAN GROUP TEST AND STANFORD MENTAL AGE WITH ABILITY IN SCHOOL SUBJECTS. MEAN OF RESULTS FOR GRADES IV TO VI

	Reading Comprehension	Reading Rate	Arithmetic	Spelling	Composite Achievement
Mean Group Test[1]	0.59	0.49	0.27	0.35	0.52
Stanford Mental Age	0.49	0.46	0.30	0.31	0.54

[1] Data for Myers Test omitted.

The mean group test stands a little higher than the Stanford on the verbal scale, according to judgments. On the scale 1.0–7.0, the Stanford is rated 4.8, the mean of the Group Tests (Myers omitted) is about 5.5. The amount of working time for the Binet is slightly greater than for the average group test, but as found in Part III, section 2, this difference would not have a great influence. Table VIII shows that the group test yields slightly higher correlations with reading and spelling, and a slightly lower correlation with arithmetic. What we get, then, is a moderate correlation with all subjects but a perfection prediction of none.

7. *Grade Differences in Correlations.*—Table X is computed from the appropriate columns of Table IV. It gives the correlations when the coefficients for all group tests (except the Myers) for all grades are averaged.

TABLE X.—SHOWING THE MEAN CORRELATIONS OF GROUP TESTS WITH

Grade	Stanford Mental Age	Reading Comprehension	Reading Rate	Arithmetic	Spelling	Composite Achievement
IV	0.38	0.59	0.43	0.30	0.47	0.54
V	0.52	0.63	0.45	0.22	0.29	0.49
VI	0.60	0.61	0.53	0.32	0.35	0.57
VII	....	0.58	0.56	0.25	0.33	0.52
VIII	....	0.50	0.43	0.25	0.33	0.47

There is a rise with the grade in the correlations of group tests with Stanford MA but the correlations of Group Tests with school subjects are about as high in the lower as in the upper grades.

If the correlations of Group Tests and the criteria of Achievement are about the same from grade to grade while the Group Tests yield a higher correlation with MA as the grade becomes higher, it will follow that the correlations of MA and Achievement will similarly go up. The data to Table XI show this to be the case.

Table XI.—Correlations of Stanford Mental Age with:

Grade	Reading composition	Reading rate	Arithmetic	Spelling	Complete achievement
IV.......	0.36	0.23	0.35	0.11	0.43
V.......	0.41	0.56	0.25	0.37	0.51
VI.......	0.69	0.69	0.30	0.45	0.67

While the increase is fairly large and quite uniform, no good reasons appear in our data to account for it.

In a preceding section it was found that the more non-verbal tests (Myers and Dearborn) showed a decrease in correlations with MA and achievement as the grade became higher, and in other sections it was found that the more verbal the material, the higher the correlation with attainment.

The verbal group tests are the same from Grade IV up and the correlations with achievement are about the same. Since the same criterion is used with the Stanford, we should look to the Stanford test itself for an explanation of the rise as the grade becomes higher. The suggestion is that the tests in the Stanford scale become *more verbal* as the MA becomes higher. To infer this from our data would be risky since unsuspected factors may enter in (for example, the older the child the more time usually required). The reader, reviewing the Stanford scale, can judge for himself.

The grade differences are important, if real. It will be worth while to devote the next section to a comparison in which the results for Grades I, II and III are included.

Part IV. Comparison of Results for Different Grades

Little weight can be given to a comparison of the results of one grade with those of another, especially where Grades I, II, and III are concerned. The measures of attainment, for one thing, are much

less reliable in the lower grades; the range of ability is larger in the lower grades since careful grading is begun with Grade IV, and finally the content of the Intelligence Tests for the primary grades is different from that of the upper grades. This is true of the Stanford as well as the Group Tests. The results are given in Table XII.

TABLE XII

Grade	1 Achievement with Mental Age	2 Achievement with Verbal	3 Achievement with Non-verbal
I	0.36	....	0.30
II	0.44	....	0.23
III	0.47	0.65	0.22
IV	0.42	0.54	0.22
V	0.51	0.49	0.17
VI	0.67	0.57	0.29
VII	....	0.52	0.08
VIII	....	0.47	−0.15

In case of the correlations of achievements with MA, the highest grades show the highest coefficients, as was pointed out in the preceding section. That Grades II and III show a larger correlation than IV is probably due in part to the fact that the range of abilities is greater in the former. The range is great in Grade I also, but the validity of the measures of achievement is small, with a resulting attenuation.

Aside from the high correlation for Grade III for reasons just mentioned, the correlations of Verbal Tests and Achievement are about the same for all grades. The non-verbal materials show a very low correlation in Grades VII and VIII but we are not certain, by any means, that the data represent the state of affairs for non-verbal materials in general, since but one Non-verbal Test (Myers) has been used in Grades IV to VIII. The Dearborn Tests, which contain both verbal and non-verbal materials, show a similar but less pronounced tendency.

For purposes of comparing one variable with another in the same grade, our data are valid. The Stanford MA and the Verbal Tests (where they are used) are clearly superior to the non-verbal. In fact, it was consistently found that the non-verbal materials added but little when the independent weights were found by the regression coeffiecient and by multiple correlation.

The Verbal Tests seem to yield higher correlations with achievement in Grades III and IV than does the MA; the coefficients are about the same for Grade V but thereafter the MA gives clearly a higher coefficient with achievement.

So far we have found two factors which influence the correlations with achievement: (1) the more verbal the material in the test the higher the correlation and (2) the longer the test, the higher the correlation, other things being as equal as we could make them. In the Verbal Group Tests both time and verbalness are equal for the Grades III to VIII. For the Stanford Test, the older the child mentally, the greater the time required. Terman's estimates of the times are:[1]

Children 6–8 years old	30–40 minutes
Children 9–12 years old	40–50 minutes
Children 12–15 years old	50–60 minutes
Adults	60–90 minutes

It is our impression that the tests become more verbal also in the higher levels. If so these two factors would account for the increasing coefficients.

Other explanations may be offered, for example, the higher levels may yield results of higher reliability. The evidence, however, is against this supposition.[2] The matter of *reliability or constancy* must not be confused with that of *validity*. It may be that the materials in the upper areas of the Terman are more valid, when constancy, time, verbalness, etc. are equal. We have no data on this point except the general finding that greater verbalness has meant greater validity when school attainment is the criterion. Another possibility is that the Stanford Test is really equally valid all the way, but that the correlation becomes higher as the pupils become more proficient—as they hit their stride—in the upper grades. The Group Tests, it might be argued, give equally high correlations all along, because they include a greater amount of reading and arithmetic and are largely measuring achievement directly. On this point we have some data. It was found, for example, that in Grade III where the pupils were rather inefficient in reading and writing at the beginning of the year, the class fell far below the norms for their age and grade in the Na-

[1] Terman, L. M.: "The Measurement of Intelligence," p. 127.

[2] Rugg, Harold and Colloton, Cecile: *Journal of Educational Psychology*, September, 1921, p. 319.

tional Intelligence Tests while their mean IQs, in the Stanford was about the same as that found for other grades—as we were expecting. The correlations of National Scores and Achievement, however, were as high as those found in other grades. These findings indicate that the Stanford Test is much less subject to the influence of school training and, if so, its general usefulness would be very much greater. It is our hope to check up this matter by comparing achievement in this and succeeding years, with the results of the various tests given in 1920.

General Summary and Conclusions

1. Other things being equal, the more verbal the material, the higher the correlation with school attainment.

(*A*) In Grades I and II, the Non-verbal Tests gave low correlations with Achievement (0.30 and 0.23, respectively) compared to 0.36 and 0.44 respectively, between the Stanford-Binet and Achievement, which is more verbal.

(*B*) In grade III, a group of Non-verbal Tests gave a mean correlation of 0.22 with achievement as compared to 0.65, the mean correlation of a group of Verbal Tests with Achievement. In this grade, the Non-verbal Tests required a longer average time than the Verbal.

(*C*) The only wholly Non-verbal Test (Myers) used in grades IV to VIII, gave much lower correlations than Verbal tests. The Dearborn Test, combining both materials, gave a higher correlation than the Myers, but a lower correlation than the mean Verbal Group Tests.

(*D*) When the materials of all tests (Grades IV to VIII) were arranged on a scale from the least to the most verbal and broken into four steps, each representing one hour teaching time, it was found that the more verbal the material the higher the correlation with the composite of achievement.

(*E*) When the individual Group Tests were arranged for the degree of verbalness, time being eliminated by the technique of partial correlations, the independent correlation (Partial r first order) with Achievement was 0.69.

2. Verbalness being equal, the greater the length of the tests, the higher the correlation with achievement.

(*A*) For Grades I and II, all tests being non-verbal, the mean correlation between length of test and magnitude of the mean correlations with all criteria was 0.69, when the SDs are made equal by use of the Rank method of correlation. Allowing the SDs to remain as they

are (Product Movement formula) the correlation is 0.49. In this case, the SDs for length (time) of tests is very large compared to the SDs for the r's.

(*B*) In Grade III, the Product Movement correlation of length and magnitude of r's with achievement is 0.76 for non-verbal tests, and 0.81 for the verbal tests.

(*C*) In the upper grades, the degree of verbalness varies so much that comparisons could be made only by the use of partial correlations. The partial correlation between achievement and time (verbalness constant) was 0.21.

3. The degree of verbalness out-weights the lengths of the test as a factor determining the correlations with achievement.

(*A*) Using the combined data for Grades IV to VIII, the following weights were obtained by the regression equation:

1. Weight of verbalness, 1.00
2. Weight (*B*) of time, 0.224

(*B*) Combining time and verbalness perfectly by the weights given above, a multiple correlation with achievement of 0.725 is obtained as compared to a partial correlation of 0.69 which verbalness alone yields, or 0.21 which time alone yields.

(*C*) The Stanford-Binet yields higher correlations in the upper grades than in the lower grades (results up to Grade VI only available). This increase is probably—but not certainly—due to (1) increasing verbalness of material in upper levels, and (2) increase in the time spent in the test.

4. When either the Stanford Test, or a verbal group test has been given, the independent contribution of the other, obtained by the regression equation, multiple or partial correlation, is not very great but probably important.

(*A*) In Grade III, the mean verbal test gives a correlation with achievement of 0.65. The addition of Stanford MA, perfectly weighted, raises the correlation (multiple r) to 0.699.

(*B*) In Grades IV, V, VI, taking mean results, the Stanford MA gives a correlation with achievement of 0.54. Adding the independent elements of the mean verbal group test, the multiple r becomes 0.605.

5. A measure of "School Attitude" obtained by judgments of teachers yields an average correlation of 0.32 with achievement, but this factor, in so far as it contributes to school success, is almost wholly included in the measures given by a combination of the Stanford-Binet and an average Group Test.

For example:

Simple r, Achievement with Stanford MA = 0.54
Multiple R, Achievement with (MA + Group Test) = 0.605.
Multiple R, Achievement with (MA + Group Test × School Attitude) = 0.611.

6. The Stanford Test and the Verbal Group Tests yield very nearly the same correlations with particular school subjects, the former correlating relatively high with arithmetic, the latter with reading and spelling.

(*A*) Moderately verbal material yields higher correlations with arithmetic than extremely verbal, but neither gives a satisfactory correlation. Extremely verbal yields higher correlations with Reading Comprehension, Reading Rate, Spelling and the Stanford-Binet.

7. The inter-correlations of school subjects are not high with the exception of Reading Comprehension with Reading Rate, which is 0.85.

(*A*) This fact suggests the need of specific tests for native aptitude for each subject.

(*B*) It raises a question with regard to the validity of the concept of the "Accomplishment Quotient" and similar practices based on the assumption of slight specialization.

(*C*) It suggests the need of correlating tests with abilities developed to the limit, rather than with abilities which are developed more or less according to the practices of the particular school.

CORRELATIONS BETWEEN BINET TESTS AND GROUP TESTS

W. T. ROOT

University of Pittsburgh

In the fall of 1920 the author supervised the testing of some 600 children in the schools of Monessen, Pa., with the Binet-Simon Tests, Stanford Revision. Early in 1921 it became possible, with the aid of Mr. Herman Gress[1] and Mr. Wade Blackburn,[1] to give a battery of mass tests to the same group. They consisted of the following: Otis Primary A (O. P. A); Otis Advance A (O. A. A); Haggerty Sigma I (H. Sigma I); Haggerty Delta I and II (H. Delta I and II); National A. I and B. I (N. A. I and N. B. I); Terman Group A (T. G. A); Mentimeter, (M.); Dearborn Series I and Series II (D. S. I and D. S. II); and Illinois I and II (Ill. I and Ill. II).

The usual statistical precautions were observed. All scores and correlations were rechecked. The greatest precautions were taken to secure uniform conditions throughout the testing. Only the splendid cooperation of the Monessen teachers made this possible. The mass tests were given weekly, on the same day at the same hour. The Dearborn Series I was given in two sittings. About 416 from the Grade I to Grade XII were given both the mass and the Binet tests. In correlating, Grades XI and XII are combined. Pearson's Product-Moment formula was used in correlating.

It has been assumed, in making the correlations, that the Binet Tests constitute the truest estimate of intelligence in so far as tests go. This may not always be the case with older (college) students as indicated in a recent article by De Camp,[2] but probably no one will take exception to the assumption that with children up to 15 or 16 years of age, the Binet Test constitutes the best single test estimate of the intelligence that can be made. Granting this, the correlation of any mass test with the Binet Test becomes of immense importance in estimating the value of the former. The administrator is anxious to know what mass test is most suitable for a particular grade or a partic-

---

[1] Mr. Gress is Superintendent of Schools, Monessen, and Mr. Blackburn is Supervisor of the grammar grades. They hope later to present a careful analysis of causes of variation in the correlations, and also an analysis of the causes leading to marked individual inconsistency in performance from test to test.

[2] De Camp, J. E.: Studies in Mental Tests. *School and Society*, Vol. XIV, pp. 253–258.

TABULATION OF CORRELATIONS

	Grade	Number	R	P. E. R
Binet with O. P. A.	1	87	0.72	0.03
Binet with O. P. A.	2	34	0.60	0.07
Binet with O. P. A.	3	36	0.63	0.07
Binet with O. P. A.	4	38	0.77	0.04
Binet with O. P. A.	all	198	0.80	0.02
Binet with O. A. A.	5	26	0.64	0.08
Binet with O. A. A.	6	32	0.46	0.09
Binet with O. A. A.	7	31	0.76	0.05
Binet with O. A. A.	8	45	0.68	0.05
Binet with O. A. A.	9	22	0.72	0.07
Binet with O. A. A.	10	25	0.55	0.09
Binet with O. A. A.	11–12	37	0.44	0.09
Binet with O. A. A.	all	218	0.80	0.02
Binet with H. Sigma I.	1	88	0.47	0.06
Binet with H. Sigma I.	2	36	0.46	0.09
Binet with H. Sigma I.	3	36	0.61	0.07
Binet with H. Sigma I.	all	160	0.74	0.02
Binet with H. Delta I.	1	88	0.71	0.04
Binet with H. Delta I.	2	36	0.28	0.10
Binet with H. Delta I.	3	37	0.57	0.07
Binet with H. Delta I.	all	162	0.76	0.02
Binet with H. Delta II.	3	36	0.62	0.07
Binet with H. Delta II.	4	40	0.69	0.06
Binet with H. Delta II.	5	25	0.58	0.09
Binet with H. Delta II.	6	32	0.60	0.08
Binet with H. Delta II.	7	31	0.82	0.04
Binet with H. Delta II.	8	44	0.79	0.06
Binet with H. Delta II.	9	22	0.44	0.12
Binet with H. Delta II.	all	232	0.84	0.01
Binet with N. A. I.	3	35	0.69	0.06
Binet with N. A. I.	4	41	0.68	0.06
Binet with N. A. I.	5	26	0.66	0.07
Binet with N. A. I.	6	32	0.72	0.06
Binet with N. A. I.	7	31	0.79	0.05
Binet with N. A. I.	8	45	0.51	0.03
Binet with N. A. I.	all	211	0.84	0.01
Binet with N. B. I.	3	35	0.67	0.06
Binet with N. B. I.	4	41	0.65	0.06
Binet with N. B. I.	5	26	0.69	0.07
Binet with N. B. I.	6	32	0.63	0.07
Binet with N. B. I.	7	31	0.67	0.07
Binet with N. B. I.	8	45	0.49	0.08
Binet with N. B. I.	all	210	0.86	0.01
Binet with T. G. A.	7	31	0.73	0.06
Binet with T. G. A.	8	45	0.65	0.06
Binet with T. G. A.	9	22	0.35	0.13
Binet with T. G. A.	10	25	0.67	0.07
Binet with T. G. A.	11–12	37	0.53	0.08
Binet with T. G. A.	all	160	0.75	0.02
Binet with M.	1	86	0.65	0.04
Binet with M.	2	35	0.49	0.09
Binet with M.	3	36	0.60	0.07
Binet with M.	4	39	0.68	0.06
Binet with M.	5	26	0.71	0.07
Binet with M.	6	32	0.53	0.09
Binet with M.	7	31	0.71	0.06
Binet with M.	8	45	0.61	0.06
Binet with M.	9	22	0.43	0.12
Binet with M.	10	25	0.68	0.07
Binet with M.	11–12	36	0.54	0.08
Binet with M.	all	407	0.88	0.01

TABULATION OF CORRELATIONS (*Continued*)

	Grade	Number	R	P. E. R
Binet with D. S. I	1	85	0.79	0.03
Binet with D. S. I	2	35	0.40	0.10
Binet with D. S. I	3	36	0.72	0.05
Binet with D. S. I	all	156	0.79	0.02
Binet with D. S. II	4	38	0.65	0.06
Binet with D. S. II	5	28	0.66	0.07
Binet with D. S. II	6	31	0.74	0.05
Binet with D. S. II	7	31	0.77	0.05
Binet with D. S. II	8	45	0.65	0.06
Binet with D. S. II	9	22	0.47	0.11
Binet with D. S. II	all	195	0.87	0.01
Binet with Ill. I	3	36	0.62	0.07
Binet with Ill. I	4	38	0.66	0.06
Binet with Ill. I	5	26	0.75	0.06
Binet with Ill. I	all	100	0.74	0.03
Binet with Ill. II	6	31	0.56	0.08
Binet with Ill. II	7	31	0.72	0.06
Binet with Ill. II	8	45	0.65	0.06
Binet with Ill. II	all	107	0.68	0.04
N. A. I with T. G. A	7	31	0.76	0.05
N. A. I with T. G. A	8	45	0.72	0.05
N. A. I with T. G. A	all	76	0.79	0.03
N. B. I with T. G. A	7	30	0.73	0.06
N. B. I with T. G. A	8	45	0.74	0.04
N. B. I with T. G. A	all	76	0.73	0.04
O. A. A. with T G. A	7	31	0.83	0.04
O. A. A. with T G. A	8	43	0.77	0.04
O. A. A. with T G. A	9	21	0.73	0.07
O. A. A. with T G. A	10	25	0.87	0.03
O. A. A. with T G. A	11–12	35	0.72	0.06
O. A. A. with T G. A	all	160	0.85	0.01
H. Delta II with T. G. A	7	31	0.86	0.03
H. Delta II with T. G. A	8	44	0.83	0.03
H. Delta II with T. G. A	9	21	0.85	0.04
H. Delta II with T. G. A	all	97	0.85	0.02
M. with T. G. A	7	29	0.79	0.05
M. with T. G. A	8	43	0.75	0.05
M. with T. G. A	9	20	0.60	0.10
M. with T. G. A	10	22	0.79	0.05
M. with T. G. A	11–12	34	0.63	0.07
M. with T. G. A	all	159	0.82	0.02
H. S. I with O. P. A	1, 2, 3	160	0.51	0.04
H. S. I with D. S. I	1, 2, 3	153	0.67	0.03
M. with O. P. A	all	188	0.93	0.01
M. with O. A. A	all	216	0.75	0.02
N. A. I with N. B. I	3–8	207	0.94	0.00
N. A. I with M	3–8	211	0.93	0.01
N. A. I with O. P. A	3–4	74	0.67	0.04
N. A. I with O. A. A	5–8	134	0.88	0.01
N. A. I with D. S. I	3	35	0.72	0.06
N. A. I with D. S. II	4–8	167	0.89	0.01
N. A. I with H. D. II	3–8	200	0.89	0.01
H. D. II with N. B. I	3–8	206	0.92	0.01
H. D. I with M	1–3	153	0.87	0.01
H. D. II with M	4–9	192	0.89	0.01
H. D. II with O. A. A	5–9	154	0.86	0.01
H. D. I with D. S. I	1–3	118	0.87	0.02
H. D. II with D. S. II	4–8	172	0.86	0.01

ular group of grades. It is hoped the accompanying correlation tabulations are a step in this direction.

The following observations may be made directly from the tabulation:

1. *Grade I.* D. S. I correlates the highest with the Binet (0.79); O. P. A next (0.72); and H. Sigma I, last (0.46). (The Sigma test it will be recalled is a reading test and not strictly an intelligence test.) The D. S. I would seem then to be the most suitable for this grade but has the disadvantage of requiring 2 days to give, owing to its extreme length.

2. *Grade II.* Judged by their correlations with the Binet Test, none of the mass tests proved satisfactory with the Grade II. The O. P. A has the highest correlation (0.59); the M., next (0.49); H. Sigma I, next (0.45); the D. S. I, next (0.40); and the H. Delta I, last (0.28).

3. *Grade III.* All of the mass tests are here more satisfactory, giving a correlation coefficient within the neighborhood of 0.60. The D. S. I is apparently most suitable (0.71); with the N. A. I a close second (0.68). The H. Delta I is least suitable (0.57). It will be noted that H. Delta II yields a higher coefficient (0.62) than H. Delta I.

4. *Grade IV.* The O. P. A, (0.77), is decidedly higher than the next most suitable test, the H. Delta II (0.69). The remaining tests, it will be noted, all lie within the 0.60s.

5. *Grade V.* The Ill. I gives the highest coefficient (0.75), with the M. a little below, (0.71). The remainder of the tests lie within the 0.60s except the H. Delta II whose coefficient falls to 0.58.

6. *Grade VI.* There is a marked difference in the correlations for this grade. The D. S. II has a coefficient of 0.74 with the N. A. I next (0.72), while the O. A. A falls lowest (0.46).

7. *Grade VII.* The highest correlation is with the H. Delta II (0.82), with the N. A. I a little lower (0.79). All of the correlations are high for this grade, lying within the 0.70s, with the exception of N. B. I (0.67).

8. *Grade VIII.* The correlations for this grade cover a wide range. H. Delta I being highest (0.79); while N. A. I (0.51) and N. B. I (0.49) are the lowest. The National correlates highly with the Binet except for this one grade.

9. *Grade IX.* The correlations are here low, and the small number of cases make the P. E.s high. The O. A. A stands highest (0.72);

then a drop to a correlation of 0.47 with D. S. II. The T. G. A stands last (0.35).

10. *Grade X.* Only three of the mass tests given cover this grade. The Mentimeter stands highest (0.68); the T. G. A a close second (0.67); and O. A. A, decidedly lower (0.54).

11. *Grades XI* and *XII.* The M. and T. G. A each give a correlation of 0.53. The O. A. A falls to 0.43. None of the tests are as satisfactory as with the lower grades.

12. *Grades I to IV inclusive.* Considering uniformity of high correlation the O. P. A seems best for these grades. It should be noted though that the Otis falls low on the Grades I and III. No test is entirely satisfactory.

13. *Grades III to VI inclusive.* It is sometimes desirable to consider these grades together. D. S. I and II make the highest and most uniform correlations; while the N. A. I and N. B. I make a close second.

14. *Grades V to VIII inclusive.* For these grades D. S. II is most desirable with H. Delta II of nearly equal value.

15. *Grades VII and VIII.* If these two grades are grouped together, H. Delta II is seemingly far superior.

16. *Grades VII to IX inclusive.* With the increase of junior high schools this grouping is now frequent. The O. A. A is apparently best.

17. *Grades IX to XII inclusive.* Grouping these grades, the O. A. A is perhaps most satisfactory.

18. *Grades VII to XII inclusive.* Considering these five grades together, there is little choice between O. A. A, T. G. A, and M. The author favors the T. G. A because it is very easy to administer, requires but 35 minutes to give, and is the simplest to score.

19. It will be noted that when the grades are pooled and correlated with the Binet, a high correlation for "all" in the tabulation is not indicative for any particular grade.

20. *Grades I to XII inclusive.* The highest general tendency to correlation with the Binet is with the Mentimeter (0.88). The Dearborn test is a close second (0.87). As these two tests cover somewhat different abilities (M. placing a premium on language ability, and D. on non-language ability) the writer suggests this combination as being the best, if two tests can be given to the entire 12 grades. The writer has been able to get more from these two series (Mentimeter and Dearborn) when he is desirous of making individual analysis than with any other combination of two mass tests.

However, if only one mass test can be given, the varied character of the Otis test makes it more valuable in analysis than either the Dearborn or the Mentimeter *alone.*

The Dearborn proved difficult to give and needs shortening and simplifying, but when this is done the author feels that this will prove to be one of our very best tests. All of the difficulties could be easily rectified. At present it is not easy for the average teacher to give, and errors in scoring are much more frequent than with other mass tests. Even as it stands, it is certainly a superior test with certain foreign children who have not yet mastered the English idiom—and this failure of mastery (with a foreign language in the home) is a bigger problem than is usually realized by those giving mass tests.

21. The correlations between the various mass tests are higher and more uniform than between the mass tests and the Binet Series. The following correlations are conspicuously high:

N. A. I with N. B. I, grades 3–8	0.94
N. A. I with M, grades 3–8	0.93
M. with O. P. A, grades 1–3	0.92
H. D. II with N. B. I, 3–8	0.92

Causes for Significant Variation in Correlations

The following are probably the chief causes for variation from test to test, and from grade to grade, in the correlations presented here:

1. Probably the greatest single factor is the difference in weight that different mass tests attach to different abilities. If a few rough captions are made, such as linguistic ability, arithmetical ability, etc., and the percentage of value attached to each caption in the various tests listed, it will be found that a marked difference exists in the relative value attached to any caption as we go from one mass test to another. It would often seem that when the maker of the test had 10 arithmetic problems *that* ability got 10 points in score; if he had 5 puzzles, puzzles scored 5; and if he happened to have on hand 20 completion sentences, completion of sentences got 20 points to the score. Be that as it may, it is certain that *chance*, rather than any *knowledge* of the relative merits of different elements in the intelligence-complex or compound, determines the *proportion* of any particular kind of psychological or pedigogical test. The *difference in proportion* is on the whole more noticeable and probably more significant as the cause of variation in score

from test to test than differences in the kind of test used by different mass–test compilers referring, of course, to the omnibus type of psychological test.

It will also be found that not only does a difference exist in the weighting of the test as a whole, but taking a certain region of the test likely to be answered, say by a Grade VI pupil, one mass test will differ radically from another both in the weight attached to different psychological captions and in the captions themselves.

2. As indicated in the line above, the mass tests differ from one another not only in the weighting of various captions but also in the actual captions included in the omnibus, or in the region of a particular grade.

3. H. Sigma correlates relatively poorly with the Binet, probably because it is essentially a reading test, and also because it demands a certain degree of reading ability. In cases where the child could read the line from the test but had attention riveted on the mechanics of reading, no action followed. Another cause of failure to respond to the test seemed to be an aversion to making a mark on the printed page.

4. It is conceivable that certain local grade conditions can play an important part; methods of teaching, predominance of certain foreign elements of a particular race, stress on certain school subjects, etc.

5. The Binet Test is largely independent of the element of time; mass tests must of necessity rest on a time basis. We do not know to what extent different subjects are benefited in one case and injured in the other, or *vice versa*.

6. Finally, marked change in the rank-order of an individual from one mass test to another or from mass test to the Binet may rest upon various individual differences. An analysis of such cases with a close study of the causes operating in an individual case, is a much needed task but beyond the scope of this preliminary report.

A METHOD OF INFERRING THE CHANGE IN A COEFFICIENT OF CORRELATION RESULTING FROM A CHANGE IN THE HETEROGENEITY OF THE GROUP

ARTHUR S. OTIS

Yonkers-on-Hudson, N. Y.

Let us suppose we know the correlation between two variables, x and y (as for example, the scores in Forms A and B of a group test of mental ability), calculated from data derived from a group of a certain heterogeneity as, for example, the pupils of a single grade, and let us suppose it is desired to know what would be the coefficient of correlation between the same variables in the case of a group of different heterogeneity as, for example, the pupils of several grades combined. The method of determining the influence of the change in heterogeneity of the group is as follows:

Let r_{xy} equal the coefficient of correlation between x and y in the first instance.

Let r'_{xy} equal the coefficient of correlation between x and y in the second instance.

Let σ_y equal the standard deviation of the y values in the first instance.

Let $\sigma_{y'}$ equal the standard deviation of the y values in the second instance.

To find r'_{xy} from r_{xy}, σ_y, and $\sigma_{y'}$, solve the formula:

$$r'_{xy} = 1 - (1 - r_{xy})\frac{\sigma^2_y}{\sigma^2_{y'}}$$

The derivation of this formula is as follows:

By the Otis difference formula[1] for correlation,

$$r_{xy} = 1 - \frac{1}{2}\frac{\sigma^2_d}{\sigma^2_y} \tag{1}$$

in which

$$d = y - \frac{\sigma_y}{\sigma_x}x. \tag{2}$$

[1] This formula was first proposed by the writer in an article entitled The Reliability of Spelling Scales, Involving a Deviation Formula for Correlation, *School and Society*, Oct. 28 to Nov. 18, 1916. It was later called the "difference formula" and the derivation shown in The Reliability of the Binet Scale and Pedagogical Scales, *Journal of Educational Research*, September, 1921. So far as the writer is aware, this formula had not been proposed by any other writer.

The quantity d is therefore the vertical distance of any point (x, y) in the correlation plot, from the line of relation the equation of which is $y = \frac{\sigma_y}{\sigma_x}x$. It is the difference, in units of the y scale, between the values of x and y for that particular case when the value of x has been transmuted into terms of y. The value of d in our suppositional case, therefore, is a measure of the amount of discrepancy between the two scores of a single individual, measured in terms of the y scale.

Now there may be, of course, a noticeable tendency for the discrepancy between the scores in the two forms of any test to vary with the magnitude of the scores. For example, if there were a tendency for the scores in the two forms to deviate less in the lower ranges, this fact would be evidenced by a pear-shaped appearance of the scatter diagram. But if the scatter diagram has a full elliptical appearance, and thus gives no suggestion of any tendency for the scores of an individual in the two forms to deviate more in one part of the scale than in another then it would be fair to assume that the tendency to deviation was the same throughout the whole range. In this case the value of σ_d would tend to be constant, and we could assume it to be constant, for all degrees of heterogeneity.

In that case,

$$r'_{xy} = 1 - \frac{1}{2}\frac{\sigma^2{}_d}{\sigma^2{}_{y'}} \tag{3}$$

in which $\sigma^2{}_d$ is the same as in equation (1).

We are now in a position to derive r' from r, knowing σ_y and σ_{y1}. Solving equation (1) for $\sigma^2{}_d$, we have

$$\sigma^2{}_d = 2(1 - r_{xy})\sigma^2{}_y \tag{4}$$

Therefore

$$r'_{xy} = 1 - \frac{1}{2}\frac{2(1 - r_{xy})\sigma^2{}_y}{\sigma^2{}_{y'}} \tag{5}$$

whence

$$r'_{xy} = 1 - (1 - r_{xy})\frac{\sigma^2{}_y}{\sigma^2{}_{y'}} \tag{6}$$

or

$$r'_{xy} = \sigma^2{}_{y'} - \sigma^2{}_y + r_{xy}\,\sigma^2{}_y$$

It must be remembered that this method does not apply to irregular scatter diagrams.

HOW THE DEARBORN INTELLIGENCE EXAMINATION STANDARDS WERE OBTAINED

WALTER F. DEARBORN

AND

EDWARD A. LINCOLN

Psycho-educational Clinic, Harvard University

The customary way of standardizing any test is to give it to as many children as possible, and combine the results to get norms which are somewhat impressive because of the large numbers upon which they are based. The theory of this procedure seems to be sound, but in practice it has given rise to some serious difficulties. Complaints have been heard that the results of some of the tests place whole classes very much too high or too low, and that the rankings obtained on two or more tests are sometimes widely different.

In the hope of obviating some of these difficulties a new method was tried in the standardizing of the Dearborn tests. The Series II examinations were given in three towns in every grade from the second through the senior class in the high school. It is hard to say just what a typical American town is, but the towns selected do not seem specialized in any way. In each of them agriculture is carried on to considerable extent, but each also does considerable manufacturing. They are large enough to support fairly large numbers of small business men, and are near enough to Boston so that the large city is a fairly open field for the inhabitants. There is in each town a fair sprinkling of children of foreign parents.

The scores were not lumped, but the results from each community were treated separately. They were distributed by months, so that it was possible to find not only the median score for the pupils of each year, but the median age as well. It has heretofore been the assumption that the children of a certain age have a median exactly at the half year, that is, for example, the children from 13.0 to 13.99 years old have a median of 13.5 years. This supposition was found to be incorrect in relation to the pupils studied for these standards. In one community the median 11 year old was only 11.33, and there were many smaller variations.

When the median scores and ages were obtained they were plotted as in the accompanying diagram. On this diagram points were chosen at each half year for standards. These were taken with the

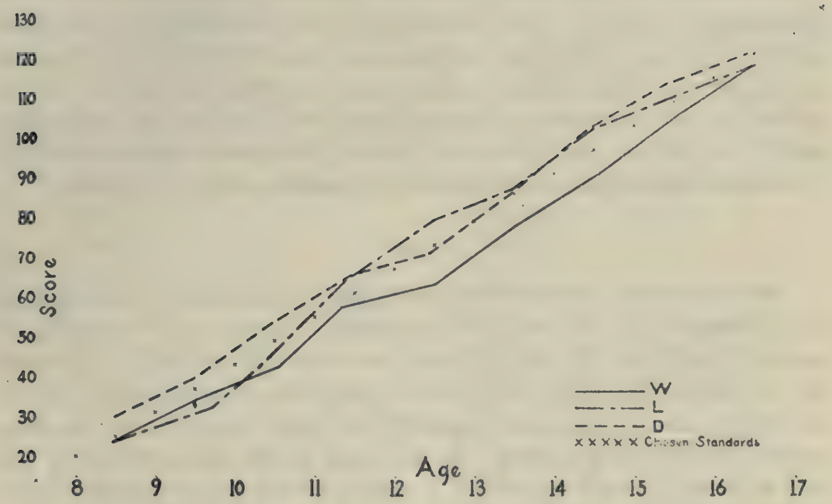
attempt to make such standards that the median child of each age should have an intelligence quotient within the normal group (0.90 to 1.10) no matter which community he was in. This criterion is very admirably fulfilled. In the twelfth year, where the discrepancies are the greatest, the median child in the lowest scoring group has an IQ of 0.93, and the median child in the highest scoring group has 1.04 for an IQ. The other deviations from 1.00 are much smaller.

It is very likely that classes, schools, and possibly one or two school systems will be found in which the distribution of intelligence quotients will be rather decidedly skewed in one way or the other.

It is believed, however, that in most of these cases the reason for the skew will be apparent. The authors have found for example, that in the foreign section of a city where the adults are engaged mostly in unskilled or semi-skilled labor the intelligence quotients on both group and individual examinations are likely to run low. It frequently appears, as may be seen on the accompanying diagram, that the pupils of a certain age or a certain grade are out of line with what seems to be the general tendency of the pupils in the community.

Series I was standardized in the same way as Series II, although it was not practicable to get results from so many upper grade children, and thus the standards from the twelfth year on had to be estimated somewhat from the continuation of the lines at their upper ends.

This method is especially valuable in that it exposes facts which are concealed when results are thrown together, and thus more intelligent treatment of the data is possible.


130
120
110
100
90
80
70
60
50
40
30
20
Score
8
9
10
11
12
13
14
15
16
17
Age
W
L
D
x x x x x Chosen Standards

THE SIGNIFICANCE OF ALPHA IN COLLEGES

CHARLES LEONARD STONE

Dartmouth College

Despite the fact that the Alpha examination was designed for purposes very unlike academic functions, much interesting material has been gathered from colleges and universities in the past two years. This statistical study, summarizing the results of the Alpha test given at Dartmouth College to 633 freshmen in the Fall of 1919 and to 622 freshmen in the Fall of 1920, is in answer to six important questions asked by college administrations: (1) How does the intelligence test correlate with total scholarship? (2) Is there a prognostic indication in the case of men separated, men on probation, and men with superior scholarship records? (3) How does the test correlate with individual subjects? (4) Is there an increasing superiority shown by the test scores as we ascend from E to A men in each subject? (5) What percentage of exceptions is there at the A and E ends of the scholarship scale? (6) Have the individual tests of the group examination any diagnostic significance with relation to specific subjects of study?

I

With the class of 1924 a definite endeavor was made, by administration influence, by an article in the college paper, and in general by campus tradition, to have all men take the test seriously. It seems safe to assume that this effort was in significant degree responsible for the higher scores, and very possibly for higher correlations, as compared with the class of 1923. In the class of 1923 the correlation with first semester grades was 0.438 ± 0.022; with second semester grades 0.333 ± 0.026; in the class of 1924 the first semester correlation with total scholarship was 0.498 ± 0.021.

II

When presented in terms of averages there is not a wide range between the men who attain a general average of B or better (173.0) and those who are separated from college for scholarship reasons (139.1). But when the distribution of these groups of scholarship is shown in terms of the intelligence quarters, the predictive significance of the test seems more hopeful.

Of the 24 men lowest in the Alpha test in the fall of 1919 (below

	Separated, per cent	Probation, per cent	*B* or better, per cent
Highest quarter	2.3	13.7	55.6
Next to highest	18.2	19.6	26.7
Next to lowest	18.2	23.5	11.1
Lowest quarter	61.4	43.1	6.7

110) twelve were eliminated within a year, whereas only five of the highest 102 (above 169) were so disposed of.

III

The following tabulation shows in the first column of figures the correlation of Alpha scores with first semester performance in the various freshman subjects, and in the second column the correlation of total scholarship with the individual subjects:

Greek	0.443 ± 0.163	0.882 ± 0.046
Latin	0.294 ± 0.047	0.842 ± 0.015
English	0.497 ± 0.021	0.712 ± 0.014
French	0.304 ± 0.030	0.739 ± 0.015
Spanish	0.119 ± 0.043	0.693 ± 0.022
German	0.363 ± 0.055	0.766 ± 0.026
Mathematics	0.379 ± 0.026	0.753 ± 0.016
Physics	0.444 ± 0.053	0.707 ± 0.032
Chemistry	0.306 ± 0.040	0.768 ± 0.017
Biology	0.220 ± 0.045	0.736 ± 0.021
Graphics	0.111 ± 0.083	0.548 ± 0.058
History	0.313 ± 0.031	0.730 ± 0.016
Physical education	0.198 ± 0.026	0.541 ± 0.019

Some of these Alpha correlations are fairly indicative. Total scholarship has much higher correlations; but of course each subject is a considerable ingredient of total scholarship, and total scholarship in the first semester has undoubtedly lower correlations with specific subjects in later college years.

We are not so vitally interested in high correlation through the middle ranges of intelligence and scholarship, however. The significance of the correlation at the extremes may well be brought out by noting what per cent of the men of each scholarship grade are found in the highest and lowest quarters of intelligence and of scholarship. The following data concern such relationships in English (highest

correlation with Alpha), Graphics (lowest correlation with Alpha), and the general average of all subjects.

	A Per cent	*B* Per cent	*C* Per cent	*D* Per cent	*E* Per cent
In Highest Quarter of Intelligence					
English	47.1	48.6	24.0	10.8	2.1
Graphics	42.9	20.0	24.1	16.7	50.0
Average	48.2	39.3	23.4	17.6	11.4
In Highest Quarter of Scholarship					
English	70.6	62.7	18.1	1.6	0.0
Graphics	71.4	27.3	17.2	0.0	0.0
Average	86.6	56.4	18.8	4.0	1.1
In Lowest Quarter of Intelligence					
English	0.0	6.5	21.4	41.7	68.8
Graphics	0.0	40.0	24.1	16.7	0.0
Average	9.0	14.0	21.8	33.1	44.0
In Lowest Quarter of Scholarship					
English	0.0	1.5	17.8	49.2	87.5
Graphics	0.0	22.7	17.2	66.7	100.0
Average	1.0	3.3	14.6	44.2	74.8

It is only fair to note that the data on English include 594 cases, that of Graphics only 64.

IV

As observed before, the averages do not represent the divergences of ability very markedly, but the following data show, on the whole, some superiority of each scholarship grade over the scholarship grade just below.

	A	*B*	*C*	*D*	*E*
English	164.3	161.7	149.9	138.9	125.4
Graphics	161.0	143.3	147.6	143.5	154.5
Total average	161.0	156.4	149.4	143.7	138.1

V

The exceptional cases are presented in terms of the per cent of *A*

men below the Alpha average of the class and of the *E* men above that average.

English	10.7 per cent *A* men	14.7 per cent *E* men
Graphics	0.0 per cent *A* men	100.0 per cent *E* men
Total average	18.2 per cent *A* men	32.8 per cent *E* men

One conspicuous objective at present should be the elimination or explanation of cases of extreme disparity between intelligence and scholarship. Probably much of this disparity can be attributed to differing degrees of motivation. One very satisfactory way to detect idlers, men too heavily loaded with extra-curricular activities, and men with unusual capacity to develop their potential, is, if we may at least tentatively trust intelligence tests, to compare intelligence percentiles with scholarship percentiles of the individual men. With such accessory data as we may get from case studies and instinct tests, a modified intelligence test will probably be one of the most valuable instruments in scientific educational administration.

VI

The very nature of Alpha, and the inclusion of certain tests of little significance—at least in their present form and degree of difficulty—make the diagnostic value of Alpha very dubious. In the rough, the language group (Greek, Latin, English, French, Spanish, and German) seems to stand out from the science group (mathematics, physics, chemistry, biology, and graphics). The data are presented in percentiles of the class of 1923.

Test	1	2	3	4	5	6	7	8	Total
Language *A* men	56.6	70.6	64.1	76.1	61.8	65.1	63.5	62.4	69.0
Science *A* men	71.4	82.7	64.1	68.9	66.8	76.1	73.9	67.0	76.1
Language *E* men	56.6	54.4	49.5	32.8	31.0	48.0	32.1	35.8	30.3
Science *E* men	56.6	51.5	53.8	40.3	31.0	41.3	29.8	38.0	31.0
Language range	0.0	16.2	14.6	43.3	30.8	17.1	31.4	26.6	38.7
Science range	18.4	31.2	10.3	28.6	35.8	34.8	43.9	29.0	45.1

Instances in specific studies, however, invalidate any definite deductions to be derived. The number completion series, which

seems slightly diagnostic of sciences, appears to be nearly as good a prognosticator of Latin ability as total Alpha. The directions test seems equally valid (or invalid) in Latin and graphics. The disarranged sentence test would seem significant in English, French, and German, but of neutral value in Spanish.

Some interesting facts emerge from this statistical accumulus. The combination of the arithmetic problems and number completion series tests correlates higher with mathematics (0.315 ± 0.027) than either test separately (0.272 ± 0.028 and 0.307 ± 0.027); but this specific combination does not correlate so high with mathematics as total Alpha (0.379 ± 0.026). On the other hand, the combination of the synonym-antonym and disarranged sentence tests correlates as well as—but no higher than—total Alpha with English ability (0.497 ± 0.021 in both cases).

All in all, the present Alpha would, from the standpoint of elective advisory purposes, seem to be as random an agent as the traditional campus method of selecting courses. But there are, nevertheless, positive tendencies which encourage the hope that a series of 10 or 12 tests may yet be evolved which will be of signal predictive value in elective advisory problems, and at the same time a partial index of prevocational aptitudes.

CHEMISTRY AND CHARACTER

THOMAS W. GALLOWAY

American Social Hygiene Association

Investigation has gone far enough to convince us that life itself, as well as the various phases and shadings of life which appear as particular functions, qualities, tendencies, and states, is largely influenced by chemical substances produced in the very act of living. For example an active living body quickly produces enough CO_2, first to accelerate respiration and in a few movements, if it is not eliminated from the system, to destroy life. The nitrogenous products of living, if not eliminated, will do the same in the course of a few days. Indeed in such a complex body as ours, we clearly have a condition in which every cell in the body pours into the blood substances which may be taken up and may modify the functions of every other cell in the body. In the evolution of this mutual adjustment of diverse tissues and their products there has arisen something like a dozen special groups of cells (*ductless glands*) whose secretions into the blood (*hormones* or *endocrines*) are known to have a special and profound influence in keeping up that balance which we call life and normality.

The researches in this most interesting field have reached the acute ink-spilling stage, and endocrines seem likely not merely to determine the fate of the individual but to activate the "fourth estate" as well.

In reviewing such books as these there are two equally tempting openings. (1) The essential biological, chemical, experimental, and therapeutic matter, which is exceedingly interesting; and (2) the implications of these for personal education and character. This review will be confined chiefly to the latter adventure,[1] although to do so is least fair to the authors, inasmuch as it is naturally in just this field that their work is most hypothetical and least satisfactory.

Bandler deals with the subject as a gynecologist, and hence emphasises the role of the internal secretions in connection with the phenomena of sex and reproduction, particularly in the female. The latter part of the book, however, discusses the instincts and emotions, mental and nervous defects, psychoses, phobias, etc. in terms of the quality and quantity of the secretions. While enthusiastic, the book is in the main reasonable.

---

[1] Bandler, S. W., M. D.: "The Endocrines." Philadelphia; W. B. Saunders Co.

Berman[1] is interesting, vivid, suggestive, picturesque, and erratic. His style—which includes the incorporation in one saturated solution (chemical!) conclusions based on experiment, on speculation, on momentum, and on temperament—is brilliant; one feels at times, unnecessarily so. A not unfair illustration of this is: "For, since menstruation is so closely connected with the phases of the moon and the tides, the rhythmicity of the posterior pituitary body may be traced to the days when the pineal was an eye in the top of the head, and in direct relation with the pituitary."

The main objection to such a mixture is not so much that imagination is introduced in such liberal proportions. This is quite legitimate; science needs imagination. Indeed, except in respect to its applications to the crass material necessities of life, one would perhaps better have imagination without facts than facts without imagination, if one must be deprived of either. Nevertheless when they are mixed, it is rather important both for the mixer and reader to know when and where, and the proportions. An adequate index to this is the great need of the book.

One has the feeling that the author himself plays a bit fast and loose with the implication of his thesis. At one place he refers to the "bubble of education," in which he is logical. And yet he recognizes the revolutionary character of "psychic conversions" (where there is no evidence of endocrine causation), in which he is right rather than logical.

It seems to the reviewer very well established (1) that there are life and death values for human beings in the endocrines; (2) that they modify growth and normality in many particulars; (3) that the inherited or acquired predominance, or the under-secretion of certain of the glands is an important factor in determining classes or types of individuals, physically and temperamentally (as, for example, the fact that the secretions of the germ cells, coupled at one time or another with certain others, make all the differences between males and females); (4) that excessive or deficient secretions can, in some cases at least, be corrected artificially, thus changing profoundly the natural personal states; (5) that these influences do extend to, and produce variations in, many at least of those personal qualities which collectively we describe as character or personality.

---

[1] Berman, Louis, M. D.: "The Glands Regulating Personality." New York: 1921, The Macmillan Co.

In a practical way, the knowledge of the endocrines will surely enable us (1) to correct many gross defects of development and functioning in matters that are basic to character; (2) to secure a better general balance of the unconscious and autonomic coordinations; (3) to diagnose native trends and types of personal balance, and by means of this to guide the individual into most suitable work and adjustments. In other words, it may well supplement our neuro-muscular and intelligence tests for vocational or other guidance. It is possible, too, that such knowledge may ultimately give us some power to increase the strength of particular traits of character, though it is at present far from evident that the endocrines are in this degree and sense "specifics."

In the opinion of the reviewer the structural and dynamic psychologist still has adequate biological grounds on which to posit the ordinary educational procedures based upon the central nervous system and its connections. Some of the grounds for this belief are: (1) There seems to be no adequate evidence that the endocrine systems or even the supposedly omnipotent "standards of the intra-visceral pressures of the vegetative system" antedate or dominate the functions of the central nervous system either in the evolution of organisms or in the development of the individual; (2) there is on the contrary abundant evidence that even the local nervous ganglia which now control these vegetative functions are made up of cells which have migrated from this central system; (3) these glands (and hence their secretions) are not the *cause* of the earliest differentiations which lay the foundation of individual development, but are rather the much later *product* of these differentiations. In other words the matter of inheritance is certainly chemical as well as physical in character, but cannot be in any strict sense endocrine—any more than it is "nervous"—in its primary nature. (4) Both nerves and endocrines are *belated* individual specializations; the endocrines—blood reactions are, with possibly one or two exceptions, entirely too slow of operation, to account for the rapid rise of the primary emotional states which accompany the sensori-motor responses of life; and hence (5) the education and controls of the individual by any form of activity and experience is still probably to be considered first and fundamentally a direct nervous (*psychological*) process, only secondarily modified, in some now unknown degree, by the resulting endocrine changes, which come largely as by-products of the nervous situation.

In estimating then the practical bearing of endocrines upon the

actual education of personality the reviewer believes that modern endocrinology is to the older theories of the bodily "humours" as modern cerebral localization is to phrenology; and that for the practical development of fairly normal people the probability is that "chemical localization" will be just about as fruitful as cerebral. The book is greatly worth a reading on the part of the discriminating educator.

ADDITIONAL RETESTS BY MEANS OF THE STANFORD REVISION OF THE BINET-SIMON TESTS

S. C. GARRISON

George Peabody College for Teachers

This article reports 468 retests by means of the Stanford revision of the Binet-Simon tests on 170 children. Of these retests, 43 were secured at an interval of 4 years, 127 at an interval of 2 years, and 298 at an interval of 1 year. In *School and Society*, June 4, 1921 the writer reported retests on 62 children at an interval of 3 years. The retests reported in that article are not included in the data presented here. Forty-three children upon whom retests were reported in that article are still in school and have been retested this year. The material secured in that retesting is included here. Goddard's revision (1911) was used originally in testing the 43 children. All other tests and retests were made with the Stanford revision.

The material presented here was secured by testing as follows: 94 cases in 1917–1918; 161 cases in 1919–1920; 157 cases in 1920–1921; and 149 cases in 1921–1922. It will be seen that the same retest is counted several times in the total of 468. For example, if a pupil was tested in 1917–1918, in 1919–1920, in 1920–1921, and in 1921–1922; we have one retest at an interval of 4 years, one at an interval of 2 years, and two at an interval of 1 year. We also have one retest at an interval of 3 years but that was included in the material reported as mentioned above. The testing was done as follows: The writer gave all the tests in 1917–1918; 138 in 1919–1920; 51 in 1920–1921; and 89 in 1921–1922. Nine advanced graduate students in educational psychology did the other testing. The students doing the testing in 1920–1921 had done little previous testing. They had, however, studied the test very thoroughly and had observed while the instructor and others gave the test. They then gave under supervision a number of tests. The students who did the work in 1919–1920 and 1921–1922 had all done previous testing and their work was carefully checked before they did any of the testing reported in this paper. All had had extensive preparation in psychology. It should also be stated that the person doing the retesting was ignorant of the results of the previous test. No comparisons between results were made until the testing program was completed.

The frequency of each age and the average difference between the results of the tests given at an interval of 1 year are shown in Table

I. The age given in the table is in every case that of the child at the second testing. There were 127 children who took the test 3 years in succession. For each of these children there are two retests and their ages are counted twice. The age at each retesting is listed.

In tabulating the ages we have listed as 12 all pupils who have passed their twelfth birthday but who have not yet reached their thirteenth.

A study of the table shows that there is some variability with respect to the average difference at the various ages. The larger average differences for the fifteenth and sixteenth year groups are probably due to the fact that there were several pupils who got practically all the tests right at both the first and second testing. These pupils would have scored higher had there been more advanced tests. We do not really know what the true IQ is for several of these pupils. Seven of the 11 pupils 8 years of age made large gains. These were all in the same grade and under the same teacher. This teacher without any previous training or practice undertook to give the Binet-Simon test to most of the children of the grade. We discovered that this had been done or was being done while we were retesting these pupils. We felt at the time that our results for this grade were influenced by that factor. If the IQs for these pupils are not included, we find an average difference for the remaining 4 of 4.2.

Table I.—Showing the Frequency of Each Age and the Average Difference Between the Two Sets of IQs for the One Year Intervals

Age	Frequency	Average difference
16	9	7.2
15	39	6.3
14	46	5.5
13	41	4.5
12	44	5.1
11	40	4.8
10	39	4.7
9	39	4.8
8	11	5.9

The differences between the IQs secured in the several testings are shown in Table II. Slightly more than 50 per cent of the differences for the 1-year-interval data lie between -2 and $+4$ inclusive. For the 2- and 4-year-interval data these limits are -3 and $+4$, and -3

and +5 respectively. Of the 468 retests, 40 (or 8.5 per cent) show a difference of more than 10. Eighty-nine per cent shows a difference of 8 or less. The table shows that there is a gain in 55 per cent of the retests and a loss in 38 per cent. The same IQ was found in 7 per cent of the cases.

Table II.—Showing Distribution of Differences in IQs Between Tests

Differences	Frequency 1-year interval	Frequency 2-year interval	Frequency 4-year interval
15	5	6	
14	2	1	1
13	3		
12	2	2	1
11	3	5	
10	2	3	1
9	3	4	2
8	7	2	3
7	12	6	2
6	16	1	3
5	20	5	3
4	23	10	4
3	27	8	3
2	22	13	2
1	17	5	1
0	22	9	
− 1	19	7	4
− 2	19	8	4
− 3	14	6	3
− 4	18	6	1
− 5	11	4	
− 6	7	5	1
− 7	11	4	1
− 8	6	2	
− 9	2		
−10	2	1	
−11	..	2	1
−12	1		
−13	2	1	
−14			
−15	..	1	1

Tables IIIa, IIIb, and IIIc give the average gain or loss and the number whose IQ was larger or smaller or remained the same at the second testing. The data are tabulated according to degree of bright-

Table IIIa.—Showing the Average Gain or Loss and the Number Gaining or Losing or Remaining the Same for the 1-year-interval Data, When Classified According to Degree of Brightness

Intelligence quotient	Average gain	Average loss	Number gaining	Number losing	Number the same
120+	1.5	...	37	21	3
110–119	1.6	...	56	32	5
100–109	0.9	...	47	32	8
90– 99	...	0.3	17	16	4
– 89	...	1.4	8	11	2

Table IIIb.—Showing the Average Gain or Loss and the Number Gaining or Losing or Remaining the Same for the 2-year-interval Data, when Classified According to Degree of Brightness

Intelligence quotient	Average gain	Average loss	Number gaining	Number losing	Number the same
120+	3.6	...	13	7	2
110–119	1.2	...	24	12	2
100–109	0.8	...	20	12	2
90– 99	0.3	...	9	8	0
– 89	...	1.3	5	8	3

Table IIIc.—Showing the Average Gain or Loss and the Number Gaining or Losing or Remaining the Same for the 4-year-interval Data, When Classified According to Degree of Brightness

Intelligence quotient	Average gain	Average loss	Number gaining	Number losing	Number the same
120+	4.0	...	7	3	0
110–119	5.1	...	8	3	0
100–109	2.0	...	6	5	0
90– 99	...	0.7	3	4	1
– 89	...	3.3	2	1	0

ness. It will be seen that a large proportion of the children test rather high. As a matter of fact the median IQ for the 170 children is 112. The high selection shown here is accounted for as follows: (1) The school is located in one of the best residential sections of the city and is patronized very largely by people engaged in the professions.

(2) A good tuition fee is charged. (3) Parents of children who do poor work are asked to withdraw their children from the school.

Several things seem to be indicated by the tables. A good majority of the children do better in the second test than they did in the first. There is a gain in 55 per cent of the cases. When the children are classified according to degree of brightness, the higher classes seem to gain more on the average than the lower. The lower classes seem to remain about the same or possibly lose a small amount. Our retests are too few in number to draw any definite conclusions on this point however. If we omit one record in our lowest group we have a small average gain showing instead of a loss in both Tables IIIb and IIIc. The large average gain shown in the higher classes in Table IIIc is doubtless due in part to the fact that Goddard's revision was used in the first testing (1917–1918). Since there does seem to be a slight gain in the higher classes, it is evident that there is a slight practise effect, that the test is relatively easier in the higher ages, or that the IQ actually increases for the higher classes. We feel that there are not enough data available yet to warrant definite conclusions.[1]

TABLE IV.—SHOWING THE RESULTS OF 468 RETESTS

	–80	85	90	95	100	105	110	115	120	125	130	135	140	145	150	155+
+155																2
150														1		3
145														3	1	1
140												2	2	2		
135												1	2	2		
130										2	3	4	2	3		
125								1	2	5	14	2				
120						2	2	11	6	9	3		2	1		
115						7	11	33	21	7	3					
110				1		17	13	21	6		1					
105				4	14	31	19	9		3	1					
100		1	1	14	21	15	4	1								
95	1	2	10	15	8	2	1									
90	1	3	10	7	2											
85	6	2	5													
– 80	14	3		1												

[1] For a summary of the data reported see Rugg and Colloton, Constancy of the Stanford-Binet IQ as Shown by Retests. *Journal of Educational Psychology,* September, 1921.

We have planned a retesting program and hope in a few years to have data which will throw light on questions raised here and elsewhere. At present we have three records on our third grade children, two on the second grade, and one on the first grade. That material is not included in this report. It is our intention to retest these children and the children in the following grades at an interval of one year until the present third grade finishes the eighth grade.

Our data are reported in Table IV. In this table the IQs from 123 to 127 inclusive are listed as 125. We found a coefficient of correlation between the tests at a 1-year interval of 0.88, at a 2-year interval of 0.91, and at a 4-year interval of 0.83.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY CECILE COLLOTON

Department of Educational Psychology, The Lincoln School of Teachers College

Intelligence Tests

A Brief History of Mental Tests. Andrew T. Wylie. Teachers College Record, 1922, January, 19–33. A very brief summary of the history and development of intelligence and educational tests. Some of the most important tests are listed with the names of the authors and the dates of publication.

Tests for Ability before College Entrance. J. B. Johnston. School and Society, 1922, Apr. 1, 345–353. Report of a study conducted at the University of Minnesota to determine the predictive value of entrance ratings of four types, (1) rank in high school classes, (2) advanced studies in high school, (3) marks on English themes, (4) score on intelligence tests. Discussion of the effect of extra curricular activities on scholarship in college.

Some Uses for Intelligence Tests. Samuel S. Brooks. Journal of Educational Research, 1922, March, 217–238. Eighth article on "Putting Standardized Tests to Practical Use in Rural Schools." Grading pupils by means of group intelligence tests supplemented by the Binet-Simon.

A Comparative Study of Four Group Scales for the Primary Grades. V. A. C. Henmon and Ruth Streitz. Journal of Educational Research, 1922, March, 185–194. Pressey's Primer Scale, Myers' Mental Measure, Dearborn's Group Test Series 1, and Haggerty's Delta 1 compared as to correlation with teachers' estimates, discriminative capacity, and conformity to natural distribution curve. Pressey, Dearborn and Haggerty of practically equal value. Pressey and Haggerty easier to administer and score. One hundred pupils in first and second grade classes tested.

The Validity of the Whipple Group Test in the Fourth and Fifth Grades. Helen Davis. Journal of Educational Research, 1922, March, 239–244. The effectiveness of the Whipple Group Tests in selecting pupils from the 4th and 5th grades for "speed classes" at Jackson, Michigan.

Does Intelligence Tell in First-grade Reading? W. W. Theisen. Elementary School Journal, 1922, March, 530–534. A study of three groups of primary pupils classified on the basis of intelligence by means of the Pressey Primer Scale. Progress of the groups in reading measured by the Haggerty Reading Test. Advantages of grouping entering pupils.

The Intelligence Testing Program of the Detroit Public Schools. Warren K. Layton. School and Society, 1922, Apr. 1, 368–372. A detailed description of the work of the Psychological Clinic of the Department of Special Education, Detroit Public Schools. The tests used; when given; uses of test results; etc.

The Value of Intelligence Tests in Universities. J. W. Bridges. School and Society, 1922, March 18, 295–302. Weaknesses of intelligence tests in colleges and universities as shown by data secured by questionnaire from 42 universities.

The South Dakota Group Intelligence Test for High Schools. W. H. Batson. School and Society, 1922, March 18, 311–315. Results of a battery of six tests designed especially for high schools and administered to 1453 students in 27 schools.

A Clinical Survey of a First Grade. Gladys G. Ide. The Psychological Clinic, 1922, January-February, 274–287. Examination of 400 first grade children by educational, psychological, and physical tests. Results of tests and recommendations on basis of results.

The Relative Efficiencies of Distributed and Concentrated Study in Memorizing. Edward S. Robinson. Journal of Experimental Psychology, 1921, October, 327–343. Two experiments conducted with students in Yale University to study various factors in the two methods of memorizing and to determine the relative merits of each. Bibliography.

Case Studies

Four Cases of Diagnostic Teaching. Gladys Poole. The Psychological Clinic, 1922, January-February, 225–229. Four case histories of children studied in the Psychological Clinic at the University of Pennsylvania. Diagnosis made on the basis of the child's response to teaching.

A Case of Special Difficulty with Reading. Bernice Leland. The Psychological Clinic, 1922, January-February, 238–244. Detailed history of a child's difficulty in reading and the remedial measures used.

Five Cases in Vocational Guidance. Rebecca E. Leaming. The Psychological Clinic, 1922, January-February, 245–255. Five case histories showing the problems met by a counselor in Junior Employment Service.

Diagnostic Problems in Educational Guidance at the Observation School, University of Pennsylvania, Summer of 1920. Gladys G. Ide. The Psychological Clinic, 1922, January-February, 265–273. Case studies of children in summer school. Need for a curriculum adapted to the "over-aged, the dull, the physically defective."

The Relation of the Conduct Difficulties of a Group of Public School Boys to their Mental Status and Home Environment. Eleanor Hope Johnson. Journal of Delinquency, 1921, November, 549–574. Report in detail of a study of 52 boys reported as "problems in conduct."

Near-Delinquents in the Public Schools. Mary Bess Henry. Journal of Delinquency, 1921, November, 529–548. Case histories of 50 children who present serious problems in the schools.

Miscellaneous

The "Double Track" System in a Small School. C. W. Odell. The Elementary School Journal, 1922, March, 544–546. Description of a flexible plan of school progress in use in a typical consolidated township school. Adaptation of state courses of study to two sections. Section A completes course in 7 years.—Section B in 8 years.

Some Data on Anatomical Age and Its Relation to Intelligence. Frances Lowell

and Herbert Woodrow. Pedagogical Seminary, 1922, March, 1–15. A study of the carpal development of 402 Minneapolis and St. Paul school children with reference to sex, chronological age, and number of permanent teeth. Comparison of carpal development with mental age as determined by the Kuhlman 1917 Revision of the Binet Test.

Child Labor and Mental Age. Raymond G. Fuller. The Pedagogical Seminary, 1922, March, 64–71. A plea for the adaptation of the school system primarily to the needs of the 85 per cent now supposedly incapable of profiting by staying in school until they are 16.

Educational Measurement as a Key to Individual Instruction and Promotions. Carleton W. Washburne. Journal of Educational Research, 1922, March, 195–206. Three necessary steps in placing a school system on an individual basis: (1) establishment of subject matter units; (2) preparation of tests completely covering each subject matter unit; (3) preparation of self-corrective practice materials. Description of work in the public schools of Winnetka, Illinois. Illustrative tests and "goals."

Short Scales for Measuring Habits of Good Citizenship. Clara Chassell, Siegried Upton and Laura Chassell. Teachers College Record, 1922, January, 52–59. Eight short scales for measuring the habits and attitudes of good citizenship are described and their derivation and construction explained. Suggestions for various uses of the results and advantages and disadvantages of the scales are given in detail.

The Description of the Performances of Pupils on Exercises of Varying Difficulty. Walter S. Monroe. School and Society, 1922, March 25, 341–343. Studies of various tests show close correlation between weighted and unweighted scores. Number of exercises done correctly practically as good a description of a pupil's performance as a weighted score.

Sectioning Classes on the Basis of Ability. C. E. Seashore. School and Society, 1922, April 1, 353–358. Description of a plan for sectioning college classes in fundamental courses on the basis of ability to progress, as shown by a competitive test at the beginning of the course. Advantages of the plan and possible objections to it are summarized.

Failures Due to Language Difficulty. Cornelia Mann. The Psychological Clinic, 1922, January-February, 230–237. Significant differences in results of testing two kindergarten groups with the Stanford Binet. Children from homes where no English is spoken at decided disadvantage in test and in first grade work.

The Effects of Practice upon the Scores and Predictive Value of the Alpha Intelligence Examination. Florence Richardson and Edward S. Robinson. Journal of Experimental Psychology, 1921, August, 300–317. Report of an experiment in administering the Alpha test to college students on three successive days. Scores on second performance probably the most reliable. Reasons for improvement.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

1. *A Mental Survey of High School Seniors.*—The idea of a mental survey of any large group of children is relatively new, but the increasing number and efficiency of group intelligence tests will naturally result in many surveys in the near future. The plea for such surveys made by the reviewer in 1918 is already bearing fruit, and they are being conducted more thoroughly and efficiently than he would have imagined possible at that time. As the significance of this sort of work becomes apparent to educators and sociologists, it will certainly lead to a great increase in the number of such surveys, because an inventory of the raw human material concerned is a necessity for a correct appreciation of every educational, social, and industrial problem. Professor Book has taken a horizontal section of the human material of the State of Indiana. The section he has chosen is narrow and very limited, but it is at the same time extremely important. Intelligence tests were given to 6188 senior high school students[1] and the results may, therefore, be considered representative of the mental caliber of senior high school students in Indiana. Only some of the most significant results can be mentioned in this review. The tremendous differences in intelligence found in different schools and in different communities is again emphasized, and the wide range of intelligence of the whole group serves again to call attention to the need for readjustment of the curriculum to the different mental levels of the pupils. Most significant for the college and the university is the fact that about as many students of inferior or mediocre intelligence are planning to go to college as students of superior intelligence. If the universities in a democracy are intended to attract and educate the youth of superior intelligence, they are failing in the sense that a large percentage of such individuals are not even planning to attend. Furthermore, the high schools themselves do not seem to be at all successful in their handling of the superior mental material, as illus-

[1] Book, W. F.: "The Intelligence of High School Seniors." Macmillan, 1922.

trated by the percentage of superior students that is retarded or held for the conventional four year course. The author rightly emphasizes again and again the waste in superior ability that the survey reveals. The failure of our educational system properly to make use of superior ability in the elementary school, the high school and the college, is gradually being revealed by intelligence tests. The relation of intelligence to the vocational choice of the pupils shows the need for vocational advice and guidance. Incidentally it should be of interest to the profession of medicine to notice the relatively low standing of the students who are planning to study medicine. The survey shows that the manufacturing districts of the state contribute a larger percentage of superior students than do the agricultural districts. The agricultural districts contribute a much larger percentage of inferior students. All districts and all economic classes and all types of schools, however, possess children of all grades of intelligence, although of course in different amounts. A slight sex difference in favor of the boys is shown, and this combined with the fact that the girls are more successful in their school work makes the author raise the question as to whether the high school is not less well adapted to boys than to girls.

The need for methods of evaluating school achievement in terms of mental ability is stressed by the author, and it is surprising to the reviewer that he has not pointed out the different ways that have already been suggested and tried out by other workers. There are many other important and valuable results in the book which cannot be mentioned in this review. It is a book that we can strongly recommend to all high school teachers and principals and it has a distinct lesson for the educator, psychologist and sociologist.

The American high school is not truly democratic, because it fails to allow for differences in intelligence, and only by so doing can it give to each full opportunity to develop to the utmost his individual capacities. R. P.

---

2. *A New Book on the Psychology of Effective Study.*—The author of this new book for teachers in training[1] has discussed the significance of training for effective study with fine insight into the underlying psychological principles. His definition of effective study sets stan-

[1] Thomas, Frank W.: "Training for Effective Study." Boston: Houghton-Mifflin Co., 1922, pp. XVIII + 251.

dards which even experienced teachers will do well to recognize. He would have teachers trained to direct pupils in the acquisition of study habits and procedures and to develop in them the ability to think and plan toward the solution of specific problems, to adopt purposes and assume the responsibility for carrying them out. He would have teachers recognize the psychology of the instincts and the fundamental considerations underlying any training which is to result in self-direction and the acquisition of socially desirable habits and tendencies. He criticises traditional practice, pointing out weaknesses and making constructive suggestions for improvement. He does this with concrete illustrations which lend clarity to the discussion, and facilitates study as he conceives it. This, together with the summaries and questions for study at the end of each chapter, recommend the book for class room use in normal schools and other teacher training institutions.

L. Z.

---

3. *A Practical Volume Based on Scientific Reading.*—The conclusions of numerous scientific studies of reading should modify current practice much more than they have. The new volume on *Silent and Oral Reading* by Clarence R. Stone[1] will certainly facilitate the adoption of scientific methods of instruction in reading. It brings together and interprets the results of psychological and educational researeh and supplies concrete and practical suggestions covering a wide range of teaching needs.

The organization of the content and the full index make it easy for teachers to use the book in the solution of specific problems. After a summary of the present situation and the outlook in Chapter I the contributions of research are discussed in the succeeding chapter. There follows a chapter on reading in the primary grades and another on the intermediate and upper grades. Four chapters are then devoted to specific problems and suggestions based on research and experimentation. Chapter IX contains a critical discussion of available reading tests and their use. The final chapter deals with individual differences and special individual and group instruction. Each chapter is followed by a group of practical problems for study and discussion. The bibliography is very brief and does not include all the references used in the text.

---

[1] Stone, Clarence R.: "Silent and Oral Reading." Boston: Houghton-Mifflin Co., 1922, pp. XVIII + 306.

We agree with Dr. Cubberley, the editor, who says in this introduction: "The contents of this volume ought to be the common property of all elementary-school principals and supervisory school officers who have supervisory oversight of elementary-school work, and be used by them as a basis for their supervision of the elementary-school work in reading. It ought also to be used by students in normal schools and teacher-training institutions in connection with the work in teaching methods and training-school practice. It would also form a very profitable study for teachers in service in connection with reading-circle study. Its simple style, absence of technical procedure, and very practical application to school room procedure all combine to make it an unusually useful book for the class room teacher to read and to follow."

L. Z.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Vol. XIII September, 1922 No. 6

AN EXPERIMENT IN LEARNING AN ABSTRACT SUBJECT[1]

EDWARD L. THORNDIKE AND C. B. UPTON

Institute of Educational Research, Teachers College, Columbia University

The purpose of the experiment was to observe the process of learning an abstract subject by adult learners in a case where the task is for them of somewhat the same novelty and difficulty as the learning of algebra is for the first-year high school pupil or as the learning of physics is for the third year pupil. We were also concerned with finding out how profitable it seemed to be for teachers to study themselves as learners in the case of such abstract material.

The subjects were a score of college graduates—teachers of mathematics. The task was to acquire understanding of certain elementary laws of electricity and magnetism. It was defined by the following instructions:

Experiment, April, 1922

Learning an Organized Abstract Subject

Study pages 1 to 22 (up to Section 15) of Franklin and Esty, "Elements of Electrical Engineering; Direct Currents" for 6 hours. Keep notes of difficulties and of how you met them.

There are six or seven copies on reference in the library.

[1] The investigation reported in this article was aided by a grant from the Commonwealth Fund.

Begin on page 1; refer back to pages XI, XII and XIII as is needed. Refer to attached notes on physics as is needed.[1] Refer to physics text-book if you find the attached notes insufficient. There are several copies on reference in the library.

Then work problems 4 to 18 inclusive on pages 448 to 453, so far as you have time in 4 hours more. Refer back to pages 1 to 22 as you need to. Keep notes as before.

The answers which the text-book provides were erased, since it seemed more instructive to make the first experiment without aid

[1] The attached notes were as follows:

C.G.S. UNITS, WHICH MAY BE NEEDED, IN SOLVING THE EXAMPLES ASSIGNED

Dyne	= force which acting on 1 gr. for 1 second increases its velocity 1 cm.
The number of dynes	= weight in grams times acceleration of gravity in centimeters per second per second.
Erg	= unit of work or energy = 1 dyne acting through a distance of 1 cm.
Joule	= 10^7 ergs.

ENGLISH UNITS

Foot-pound	= force required to move 1 lb. a distance of 1 ft.
Pound-inch	= force required to move 1 lb. 1 in.

DEFINITIONS

Torque equals the force times the distance through which the power works.

Center of gravity is the point on which any object will balance.

Armature, a core of metal surrounded by a coil of wire, rotating near the poles of magnets in a dynamo.

Dynamo, a machine to convert mechanical energy expended upon it into electrical energy.

Equivalent units of measure that may be needed in solving the examples assigned:

1 horse power	=	33,000 ft.-lbs. per minute
1 horse power	=	746 watts
1 watt	=	0.00134 horse power
1 joule	=	0.74 ft.-lb.
1 watt	=	44.3 ft.-lb. per minute
1 cm.	=	0.3937 in.
1 ft.	=	30.48 cm.
1 lb.	=	453.6 gr.
1 gram	=	0.0353 oz. avoirdupois
1 kilogram	=	2.2046 lb. avoirdupois

from them. We hope to repeat the experiment with a similar group using the answers.

It was explained to most of the subjects that the 6 hours of work on the text, the 4 hours of work on the text, and the problems could be divided up and alternated in any way that the subject might choose. All work was signed with a number or pseudonym. The subjects were asked to hand in all their work and to put down in writing any matters concerning the learning process, such as difficulties and the means taken to overcome them which seemed instructive, taking out time therefor.

This last injunction resulted in only rather meager notes, probably for two reasons. First, the teacher does not, unless he has an exceptional interest in the psychology of learning, watch himself learn when he learns. On the contrary he becomes absorbed in the content which he is trying to master. In the second place, there is a natural tendency to avoid the labor of making such notes.

The 10 hours were spent at times of the subject's choice during a week. At the close of the week certain questions about their experiences were asked of all the subjects and a 12-minute test was given. What follows in this article is the writer's opinion based on the answers to these questions, the notes on learning, the written work, and the test.

The experiment seems a useful one as a means of increasing the subject's appreciation of pupils' difficulties and sympathy with their efforts, and as a means of increasing the experimenter's knowledge of learning. The task assigned is somewhere nearly as hard for college graduates who know physics only from an elementary course, as beginning algebra and physics and economics are for their students. This is evidenced by the records with the problems. The highest score was 12½ correct out of 15; the lowest was 0; the median was 5; half were from 3 to 7 inclusive. The errors gave by their nature, and especially their variety, much the same impression of inadequacy, confusion and carelessness that one gets from the results of a hard assignment to a high school class in an abstract subject.

It is not grossly inaccurate to say to a group of teachers who have done this experiment: "This is to you approximately what a hard series of lessons is to a high school class. The difficulties and discouragements and confusion which you have felt are what they feel. Although you did your best, your work looks stupid and careless, much as theirs does. It looks in places as if you 'did not try' or

'would not make yourself think,' or 'did not keep your attention on what you were doing.'" The experiment should be a warning and protection against underestimating pupils' difficulties and imputing to perverseness or lack of effort results which are really due to the general laws of mental action. It should enforce the general lesson of psychology that commands and exhortations and rebukes are only a very small part of teaching, the great part being to discern the forces of the pupils' own minds and hearts and manœuver them to desirable ends.

So much for the benefits to be derived from the experiment in making teachers more humane and reasonable. With respect to our knowledge of the learning process itself, its chief result is its evidence that gifted, highly trained adults are much more like ordinary untrained children in methods and procedure in learning than it has been the fashion to suppose. It has been customary to contrast the inferior young, untrained child sharply with the superior, trained adult in that the former is impulsive, leaping before he looks; is uncritical, accepting any idea that happens to strike him regardless of its appropriateness; and is confused and dreamy and irrelevant, lacking clear-cut ideas of what he is to think and why he is to think it. This contrast is treated as general, pervading all the thinking of the two types of mind, and constitutional, depending on their distinctive natures, not on the tasks in which they engage. The statement that superior trained adults seem more rational, critical and clear-headed in large measure because they do only what is easy for them, would be regarded by all educational psychologists as a paradox and by most of them as a very silly one. Yet our inspection of the work of this score of superior trained adult thinkers strongly suggests that the paradox has a large element of truth, that the difference in thinking is not all-pervasive and constitutional, but is, in part, a consequence of specialized habits.

Consider, for example, the answers given by these gifted trained adults to the question "Why are magnets which are used to pick up pieces of iron bent into U form?" which was asked at the end of the 10 hours of study of the elementary principles of magnetism. Many of them show an impulsive, uncritical, and confused thinking to a notable degree—to the person who is really master of the topic.

(*A*) The intensity of magnetic field is increased by bringing poles near each other. Attraction between two poles is product of their intensities divided by distance squared, and as distance is decreased the intensity increases inversely as its square.

(*B*) Magnetic force tends to concentrate at one spot or another. Therefore, if the iron is bent horseshoe fashion and magnetized, there is more strength.

(*C*) To get both poles on a level to attract either or both poles of the article you want to pick up.

(*D*) In order that the object may lie along the lines of intensest force, thus the strongest pull being conducted through the material picked up.

(*E*) Magnets are bent into U form in order to create a magnetic field of . . . .

(*F*) Magnets are bent into U form to increase intensity of field. The two poles being . . . .

(*G*) So that the poles are more convenient for use. More handy for use than a long magnet.

(*H*) The magnet is in U form to bring poles nearer together.

(*I*) To strengthen field; $F = \frac{m'm''}{r^2}$ as distance r decreases, F increases.

(*J*) Pull is both ways and stronger north and south poles give equal pull.

(*K*) Bringing both poles together condenses the external magnetic field, thus making it stronger. Both poles are brought to bear on an object, thus . . . .

(*L*) To attract both north and south pole. The flux is from north to south pole; by having U-shaped magnet the two poles are near enough to make the circuit.

(*M*) So that both poles may operate at once; putting the poles together concentrates the field, *i.e.*, shortens and pushes closer the lines of force.

(*N*) Because the closer the poles are together the more intense is the magnetic flux.

(*O*) Magnets are bent into U form to bring the poles nearer together.

(*P*) Bent to give two poles so that the current is attracted to the second pole, otherwise there would be no flow of current.

(*Q*) U form is due to the fact that like poles repel and unlike poles attract. If the poles are close together the force of attraction is greater and exerts greater force on external objects.

(*R*) The poles are opposite in attraction and have power that is concentrated.

(*S*) So that the entire force of the magnet is at one point and so that both directions of force may be exerted at once.

(*T*) Double the strength, or attraction on the iron.

Impulsiveness and uncriticalness are witnessed further by the large number of wrong answers in comparison with the number of omissions or confessions of inability. In the book problems there were (counting each task requiring a separate answer as one problem) 341 answers given of which only 53 per cent were right. In the case

of few of these was there any apology or other expression of doubt about the answer. In the test problems a similar state of affairs obtains. The extreme of uncriticalness appears where the thinker merely "fishes about" for any promising formula and puts the number of the problem into that formula, almost haphazard. Consider these answers all given to the same problem and all wrong.

15,503 lines
15.5035
93,020 lines
348,000 cos 63
174,000 cos 63
4.739 gauss
78,994 dynes
7.8996

A general mental confusion is evidenced by much of the work, especially perhaps by the very frequent failure to define the answer numbers as dynes, gausses, maxwells, lines, centimeters, or whatever they should be, and by the frequent attachment of erroneous names to the answer numbers.

Thus in the test the question: "What is the intensity of the magnetic field due to a unit pole at a distance of .001 cm.?" received a correct number answer six times but the name was correct only once.

The "childishness" of these adults appears also in their eagerness to learn from some authority whether the answers which they obtain are right, rather than trust to their own proofs and checks.

We are not, of course, saying that the work of first-year high school pupils *at this task* would not be easily distinguished from these adults' work at it. Nor do we say that these adults are as childish at it as first-year pupils are at their tasks in mathematics, grammar, or science. What is claimed here is not that the difference between the thinking of superior trained adults and that of children is zero, but that it is less than the orthodox educational psychology and child psychology of the last two decades has taught. In proportion as a task for thought is novel and hard, the superior trained adult tends to jump at conclusions instead of planfully mastering each necessary step, to let results stand without surety that they are true and useful, and to become confused, making computations and statements without any clear realization of what he is doing and why he does it.[1]

---

[1] I may be permitted to note that I observed these tendencies in my own work at this task.

On the Psychology of Teaching in General

The experiment provides evidence corroborating certain pedagogical doctrines which are already accepted or on the road to acceptance. There was general agreement among the subjects that each item of fact or principle should be applied as soon as learned, that miscellaneous problems requiring the selection of appropriate principles should be given later, and that many short problems with a minimum of computation and interpretation of complex situations should be given first. There was general agreement that, even with the very scanty problem list of the book used, the problems explained the text as truly as the text explained the problems. There was substantial agreement that the learner should be enabled to ascertain whether his work was correct very soon after he finished any part of it and very often along its course. Sample problems requiring the use of formulas should be solved in the text as illustrations of the formulas.

The irritation caused by features of the task that were, or at least seemed, irrelevant, such as laborious computations, a problem about the torque on a pulley, and the recalculation of an answer in another system of units, was notable. Irritation at being thwarted by a difficulty that was, or at least seemed, not provided for by the text was also great. The text left, or seemed to leave, the meaning of gauss and the relation of a gauss to the force exerted by a unit pole at a distance of 1 cm. to be inferred. We estimate that if it had stated it clearly and emphatically there would have been a saving per person of 30 minutes time, much irritation and at least ten per cent in errors.[1]

Finally the facts of the experiment suggested to the writer as highly probable an important amendment to the common view of the difference between the "easy" subjects like English, History, French, or book-keeping and the "hard" subjects like mathematics, Latin, or physics. The common view is that the latter are hard because they are more abstract, more organized, more rigorous and more precise. They require, to a greater degree, analysis and reasoning in place of memory, systematized related knowledge in place of an accumulation of details, an absolute adherence to certain rules of the game rather than a general following of their spirit, and 100 per cent precision in the operation of the mental bonds instead of a certain free play for

[1] Neither this paragraph, nor anything else in this report, is intended or should be interpreted as an adverse criticism of the text-book pages used. Our experiment put them to a use very different from that for which they were designed.

errors of moderate amount. This is all true; but it may also be true that part of the greater difficulty is due to the sheer unfamiliarity of the date, the greater amount to be learned per unit of time, the more subtle symbolism, and the larger number of missing links—facts that have to be read between the lines, connections that have to be worked out by the pupil by inference. If high school pupils studied the history of philosophical systems or of credit facilities, history would be made harder by the unfamiliar data. If, when they had studied a certain phenomena, say the causes of revolt in Greece and Rome, they were expected to apply that knowledge to all cases of revolt anywhere, distinguishing when it was and when it was not applicable, history would be harder because learning the causes of revolt in Greece and Rome would then be learning a great deal more than a chronicle plus certain suggestions. If, in the text-book on history, *e* stood for envy, *r* for race antagonism, *p* for George Washington, the area of a square for the number of people concerned, l_1, l_2, l_3 etc., for certain typical cases previously taught, and the like, history would be harder. If any facts that a competent thinker could infer were omitted from statement in history (as if we read "Charles" instead of "King Charles," "Gregory" instead of "Pope Gregory." "He infringed upon the six most important civil rights of the population," instead of a long paragraph stating in detail what he did, history would be harder.

In the experiment all these four factors appeared: (1) Strength of pole, magnetic field, intensity of magnetic field, and magnetic flux are hard partly because they are strange to us. (2) Formulas like $F = mH$ or $\Phi = sH$ look like small bits of learning, but each is really, if understood, a large body of fact and a still larger possibility of handling other facts. Thirteen such formulas in 6 hours means a very large amount of learning per hour. (3) The symbolism which will in the end economize thought is in the beginning a burden. (4) Consider the number of inferences that a student must make to obtain a true and adequate learning of this single half page.

> "*Magnetic Flux.*—Consider a plane surface, *s* square centimeters in area, stretched across and at right angles to a uniform magnetic field of intensity H. The product sH is called the *magnetic flux* across the surface. That is:
>
> $$\phi = sH \qquad (4a)$$
>
> in which Φ is the magnetic flux across a plane surface of area *s* square centimeters at right angles to a uniform magnetic

field of intensity H. When the plane surface is not at right angles to the uniform magnetic field then:

$$\phi = sH \cos \theta \qquad (4b)$$

iu which θ is the angle between H and the normal to the surface. When the field is not uniform or when the surface is curved then:

$$\Delta\phi = H \cos \theta . \Delta s \qquad (4c)$$

in which $\Delta\phi$ is the flux across an element of the surface of which the area is Δs, H is the intensity of the field at the element of surface, and θ is the angle between H and the normal to the element of surface. In this case the total flux across a finite surface is found by integrating equation (4c) over the finite surface."

So, we repeat, the experiment suggests that these other characteristics deserve consideration along with abstractness, organization, rigor and precision. Precision and rigor may be rated as indispensable features of secondary education. Abstractness, organization, and subtle symbolisms are very desirable features of it for those whose intellects can handle them. What I have called *reading between the lines*, supplying necessary interpretations and connections, is desirable in moderation. The mere difference in amount of learning whereby an assignment in English or history seems more than it is and an assignment in mathematics or physics is much more than it seems, is not of any intrinsic educational value. The difference in strangeness is perhaps of negative value educationally, it being unwise to try to understand, analyze, compare and reason about facts until we have a certain familiarity with them as facts. In so far as Latin is hard because of the sheer strangeness of datives and ablatives, its hardness is chiefly an unfortunate interference to be overcome. In so far as algebra is hard because the use of letters to represent numbers is strange, the lesson to its teachers is to remove that much of its hardness as quickly as they can by suitable experiences with formulas and their evaluation.

A CLASS EXPERIMENT IN LEARNING

WALTER F. DEARBORN AND EDWARD A. LINCOLN

Psycho-educational Clinic, Graduate School of Education, Harvard University

In presenting to classes in Educational Psychology the facts about learning, the effects of practice and transfer of training it seems highly desirable to adopt a method which will furnish to the students evidence of an objective nature. The limited time at the disposal of the lecturer in the ordinary introductory course in this field precludes the possibility of adapting the typical experiments in learning to fit the occasion. The studies of Bryan and Harter, Swift, Book, and others have dealt with the learning of telegraphy, typewriting, or some other subject matter which required lengthy practice extending sometimes over months. In a previous article by one of the present writers a preliminary statement of the problem was made and suggestions were offered for experiments which seemed to meet the requirements of the situation.[1] Several years later there was published a more elaborate scheme for carrying out a class experiment in learning.[2] This experiment did not prove entirely satisfactory, but it has since been modified, after trial in a number of classes, and it now seems to serve its purpose excellently. It is the object of this paper to report the methods and results of the experiment as given in its final form.

The class in Educational Psychology was divided arbitrarily according to the seating arrangements into two groups of approximately equal size. On the first day of the experiment all members of the class were given three tests, hereafter designated as the Test Series. The first of these is called the Digit Letter Substitution Test. Sheets of paper were distributed, each containing at the top circles in which were inserted the numbers from 1 to 26. Each circle also contained one of the letters of the alphabet, the distribution being a random one. Under the code were printed a number of sentences, and the task consisted in translating letters into digits for a period of 5 minutes. As shown in Fig. 1 the arrangement of the sheet permits of rapid scoring. The subjects were told to work for accuracy rather

[1] Dearborn, W. F.: Experiments in Learning. *Journal of Educational Psychology*, Vol. I, pp. 373–384.

[2] Dearborn, W. F., and Brewer, John M.: Methods and Results of a Class Experiment in Learning. *Journal of Educational Psychology*, Vol. IX, No. 2, Feb., 1918, pp. 63–82.

O	V	C	F	I	K	B	M	
1	2	3	4	5	6	7	8	
T	X	E	W	N	H	G	L	D
9	10	11	12	13	14	15	16	17
J	S	R	U	Z	P	Y	A	Q
18	19	20	21	22	23	24	25	26

in the wind. The boys also	5	13	9	14	11	12	5	13	17	9	14	11	7	7	24	25	16	19	1	
spin tops, and they play	19	23	5	13	9	1	23	19	25	13	17	9	14	11	24	23	16	25	24	
"seesaw," and jump the rope.	19	11	18	~~25~~ 19	25	12	25	13	17	18	21	8	23	9	14	11	20	1	23	11
Boys in Korea are fond of	7	1	24	19	5	13	6	1	20	11	25	25	20	11	4	1	13	17	1	4
fishing. Nearly every																				
boy has a fishing rod and																				
goes fishing whenever																				
he can. Sometimes the																				

FIG. 1.—Sample of digit-letter substitution test.

than speed, and to correct mistakes, and no account was taken of the errors in making up the scores.[1]

The second test is designated as the Complex Dotting Test. For

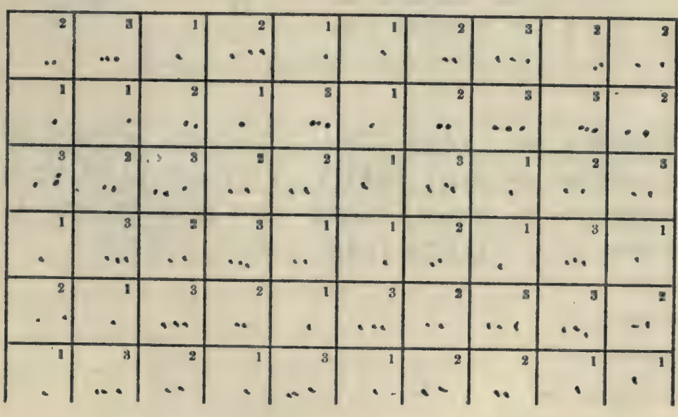


FIG. 2.—Sample of dotting test.

[1] In one of the previous experiments the curve of errors was plotted, and it was shown that the errors might be disregarded.

this test sheets of paper are ruled into half-inch squares, each of which contains either the number 1, 2 or 3 (See Fig. 2). The numbers are distributed at random throughout the 150 squares on the sheet. The task is to put in each square as rapidly as possible the number of dots called for by the number in that square. One minute is allowed, and the score is the number of squares completed.

Test 3 is a modification of the code test of the Stanford-Binet Scale.[1] This code is reproduced in Fig. 3, and Fig. 4 shows part of a test sheet. Five minutes are allowed for this test, and the score is the number of letters translated. As in the other tests, no allowance is made for errors.

After these tests had been given the class was practiced for 10 minutes in the substitution of shorthand symbols for about a hundred of the most common words and word phrases in the English language. These symbols were printed on a key, a copy of which was given to each subject (Fig. 5). The material in which the substitutions were made was furnished by double spacing a book which was being printed at the University Press at the time and striking off sufficient copies on inexpensive paper. Figure 6 shows a section of this practice material in which the symbols have been substituted.

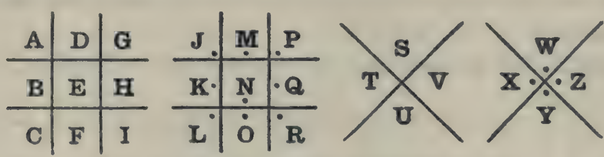


FIG. 3.—Code for code substitution test.

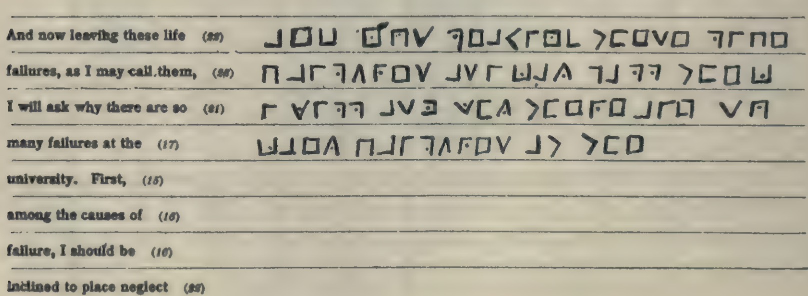


FIG. 4.—Sample of code substitution test.

[1] This code test seems to have first been described by Healy and Fernald in *Psychological Review Monographs*, Test No. XI, Vol. XIII, No. 2.

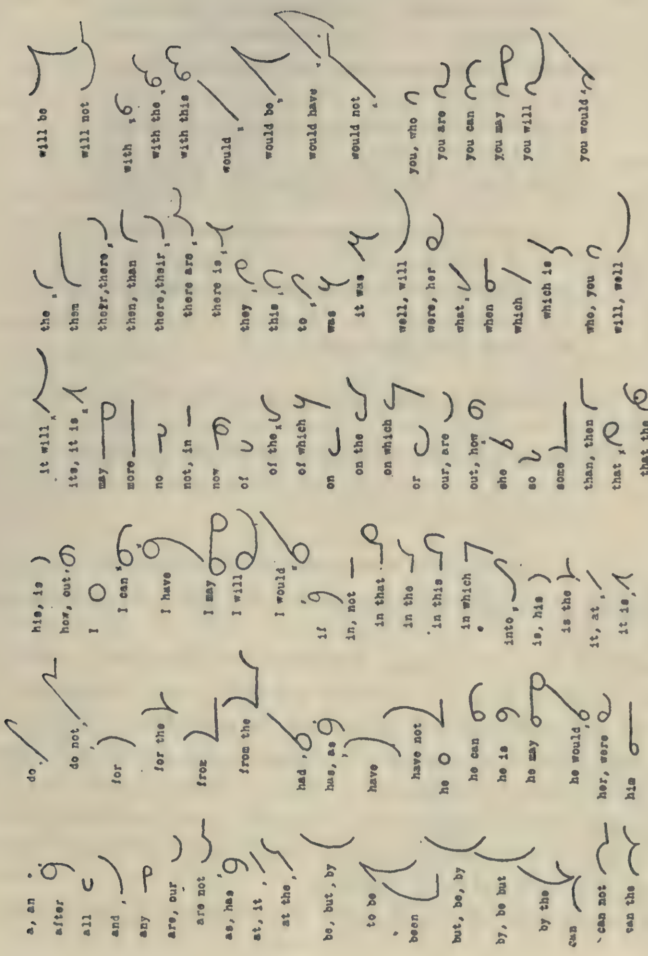
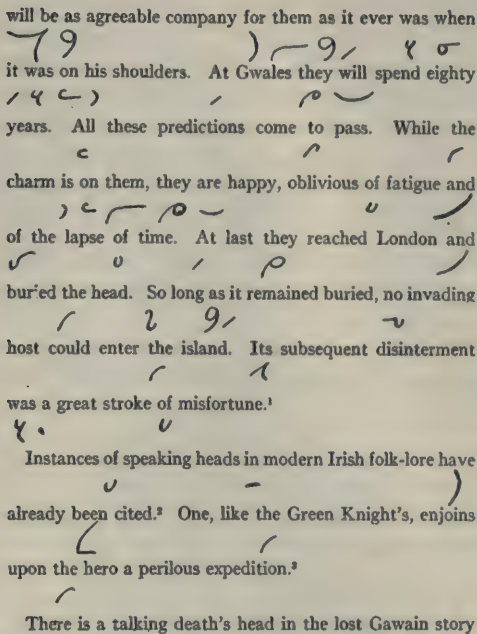


FIG. 5.—Shorthand key.


will be as agreeable company for them as it ever was when it was on his shoulders. At Gwales they will spend eighty years. All these predictions come to pass. While the charm is on them, they are happy, oblivious of fatigue and of the lapse of time. At last they reached London and buried the head. So long as it remained buried, no invading host could enter the island. Its subsequent disinterment was a great stroke of misfortune.[1]

Instances of speaking heads in modern Irish folk-lore have already been cited.[2] One, like the Green Knight's, enjoins upon the hero a perilous expedition.[3]

There is a talking death's head in the lost Gawain story

FIG. 6.—Sample of the practice material showing the shorthand substitutions.

During the next 15 days, exclusive of Sundays, this practice continued. One group of the class practiced one 20-minute period a day, and the other group practiced 20 minutes in two 10-minute periods, one in the morning and the other in the evening. Each man counted and recorded the number of substitutions made in each period, also noting the mental and physical conditions under which he did his work. At the end of the experiment each man tabulated his results and plotted his own curve of learning.

On the fifth and tenth days of the practice period check tests

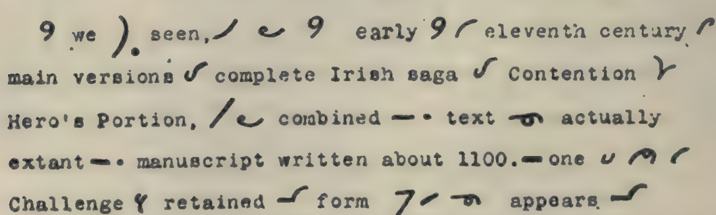

9 we). seen, 9 early 9 eleventh century main versions complete Irish saga Contention Hero's Portion, combined — • text actually extant — • manuscript written about 1100. — one Challenge retained form appears

FIG. 7.—Sample of check test material.

were given in the class. These tests reversed the operations of the practice. In a mimeographed passage from the same book which was used as practice material, shorthand symbols had been substituted in place of the words and phrases which appeared in the key, and under each symbol the proper word or phrase was written by the subject. Four minutes were allowed for each of these tests. In Fig. 7 a section of one of these tests is reproduced.

The final period of practice was held in the class room after the students had been working individually for 14 days.[1] At this time the test series was repeated.

As it was desired to show the ideal procedure for a transfer experiment the test series was given to a control group on the same days in which the practiced group took their first and last tests. This control group was composed of a class in Educational Psychology at Radcliffe College and a class in the History of Education at Harvard. Each class contained both graduate and undergraduate students. The following diagram, patterned after one in the previous article, illustrates the general plan of the experiment.

An Experiment in Learning

Practiced Group		
1st day	2nd, 3rd..................15th day	16th day
3 tests and first practice given	Practice series with shorthand symbols	Last practice given; 3 tests repeated.
Unpracticed or Control Group		
1st day	2nd, 3rd..................15th day	16th day
3 tests given	No intervening practice	3 tests repeated

After each student had made his own table of results and had plotted his curve of learning the records were collected and turned over to various groups in the class for the working out of certain problems which were made the subject of class reports. The findings presented in the more important of these reports are set forth in the following paragraphs.

Perhaps the greatest significance of the results of this experiment lies in the light they shed upon the problem of the transfer of training. The average scores of both practiced and control groups in the initial and final tests are set down in Table 1, together with the percentages of improvement shown in the test and practice series.

[1] The length of the practice may be changed to suit the conditions of the course in which the experiment is used.

TABLE I.—SCORES AND PERCENTAGES OF IMPROVEMENT OF PRACTICED AND CONTROL GROUPS

	Average initial score	Average final score	Per cent improvement
I. Practiced group:			
(*a*) Shorthand Substitution	72	467	548
(*b*) Digit letter Substitution	98	116	18
(*c*) Dotting	102	118	16
(*d*) Code Substitution	103	118	15
II. Control group:			
(*a*) Shorthand Substitution	No practice given.		
(*b*) Digit letter Substitution	93	103	11
(*c*) Dotting	102	113	11
(*d*) Code Substitution	92	116	26

There was plainly a very great amount of improvement in the ability to make the shorthand substitutions, as the class made an average gain of 548 per cent. The gains in the test series, however, were much smaller, namely, 15, 16, and 18 per cent. These test series gains become even smaller when the results of the control group have been taken into consideration. In the Digit Letter Substitution the mere repeating of the test in the control group shows an 11 per cent improvement. As the practiced group only improved 18 per cent in this test there is left a net improvement of 18 minus 11 or 7 per cent which may be attributed to transfer. In the same way we find a 5 per cent improvement which may be attributed to transfer in the Complex Dotting test, and in the Code Substitution there is no transfer or facilitation at all, but rather interference, as the control group improves 26 per cent while 15 per cent is all that is found in the practiced group.

In comparison with the improvement of 548 per cent which took place as a result of the practice it seems fair to conclude that the transfer was insignificant. We must, however, enter into another phase of the problem. Before we can decide on the significance of the above percentages we must know something about the zero points and the physiological limits in the various tests. Fortunately, the Digit Letter and Code Substitution Tests had each been used for practice material in previous experiments. In 12 days of practice

a group of students improved 463 per cent in the Digit Letter Test, and another class improved 537 per cent in Code Substitution.[1] Thus it seems that the possibilities of improvement in these tests were considerable. The fact that so little of this possible improvement appears in one test and that there is not an improvement but a loss in the other leads us to the conclusion that the skill acquired in one field was not transferable to another field, even though the second was very similar to the first. The possibility of one sort of learning interfering with another is also neatly illustrated.

The experiment also furnishes excellent material for the study of the effects of practice on the variability of a group. To get data on this point the average deviation of each group was calculated every third day throughout the experiment. These average deviations, together with the averages and coefficients of variability on the same days are tabulated in Table 2. On Plates I and II the deviations are shown by vertical lines.

TABLE II.—AVERAGES AND AVERAGE DEVIATIONS IN SHORTHAND SUBSTITUTION PRACTICE

Day	Group I		Coefficient of variability	Group II		Coefficient of variability
1	Average	150	0.08	Average	168	0.12
	Average deviation	12		Average deviation	20	
4	Average	344	0.13	Average	360	0.09
	Average deviation	46		Average deviation	34	
7	Average	535	0.12	Average	507	0.11
	Average deviation	62		Average deviation	57	
10	Average	665	0.13	Average	636	0.15
	Average deviation	88		Average deviation	84	
13	Average	844	0.11	Average	814	0.11
	Average deviation	90		Average deviation	90	
16	Average	960	0.11	Average	1001	0.11
	Average deviation	89		Average deviation	106	

The figures indicate that the heterogeneity of both groups was much increased by the practice. In Group 1 the average deviation on the first day was only 12, while on the last day it was 89. In Group II there was a rise from 20 to 106 in the average deviation.

[1] See above reference.

Manifestly the ability was more widely distributed at the end of the practice than it was at the beginning. This is shown in a rather striking way on the plates by the lines showing the daily individual performances of the slowest and most rapid learners in each group. When this change is expressed in terms of the coefficient of variability, it is seen that the increase in variability bears a fairly constant relation to the increase in the average skill of the groups.

The division of the class into two groups was for the purpose of investigating the effects of different distributions of practice time. The results of this study are best shown in Plate III where the daily averages of the two groups are plotted. These averages are tabulated in Table III. It will be seen that there are not large differences in the two curves, but the weight of evidence seems slightly in favor of the two short periods a day rather than one long one. The Group II curve is very smooth, while the daily performances of Group I

Table III.—Daily Average Scores of Two Practiced Groups

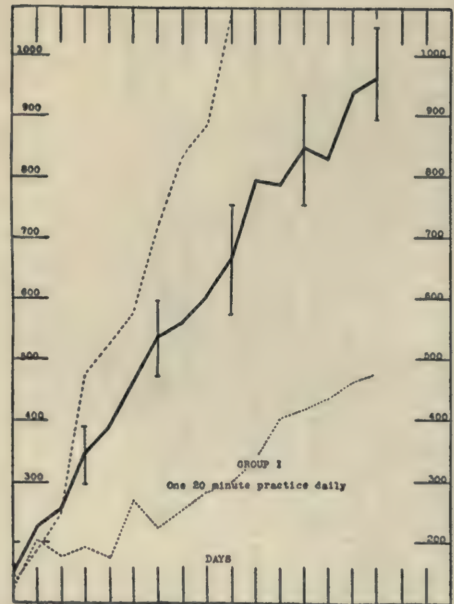
Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Group I	150	223	254	344	390	462	535	560	602	665	793	786	844	829	936	960
Group II	168	247	304	360	414	446	507	545	588	636	686	740	814	859	908	1001

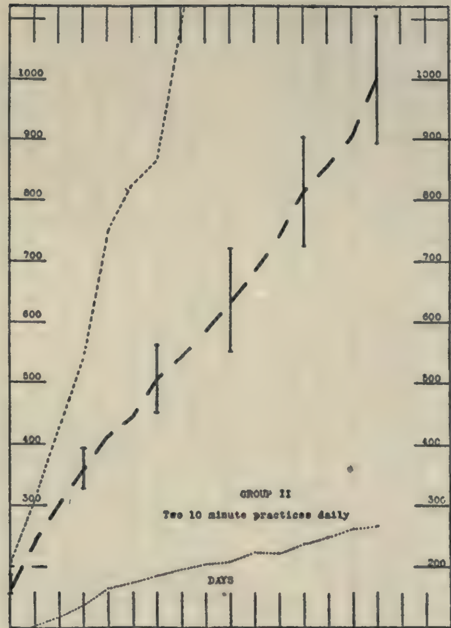
show more fluctuation, especially from the tenth day on. There are two days when distinct losses are registered in the one period group, and these would be very undesirable in any learning because of the bad effect on the attitude of the learner toward his work when he fails to make progress.

The experiment is useful in a number of other ways. Plotting of the individual records, for instance, gives excellent illustrative material for the study of individual differences. In the usual class there will probably be found curves which will illustrate all the common characteristics of practice curves such as plateaus, daily fluctuations, and so on.

If the subjects keep careful records of the conditions under which they work at different times during the practice period some interesting facts may be obtained concerning the effect of various factors on the performances of individuals. Changes resulting from interest, fatigue, distractions, time of day when practice is done, and like circumstances are likely to appear, and add greatly to the value and interest of the work.

The scores in the various tests furnish excellent material for illustration of the mathematical problems of measurement. The





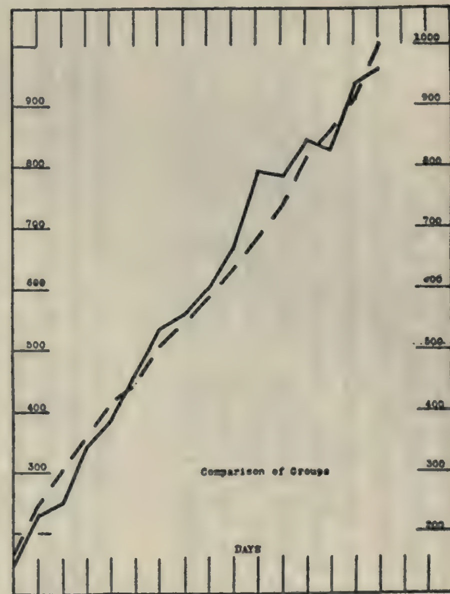


Plate I.—Daily averages of Group I, with best and poorest records of the group. The perpendicular lines every third day show the average deviations on those days.

Plate II.—Daily averages of Group II, with best and poorest records of the group. The perpendicular lines every third day show the average deviations on those days.

Plate III.—Comparison of the two groups to determine the better time distribution in practice periods.

data obtained proved especially useful in illustrating to the class the methods of numerical and graphic correlation.

While it is believed that the findings will ordinarily be in agreement with the more general and more extensive experiments in the same field, it is possible that there will be differences in some of the select groups which make up college and normal school classes. Care should be taken, therefore, to make it clear, when this or a similar experiment is used, that the results do not prove or disprove the various educational theories concerning which they furnish evidence.[1]

---

[1] This experiment was originally begun as a contribution to the work of a committee of the American Psychological Association on Class Experiments in Psychology. If there is sufficient demand, the materials will be printed in quantities and supplied at cost to those interested.

LANGUAGE ERROR TESTS

G. M. WILSON

Boston University

For 10 years the writer has been interested in language errors and for the last 5 years has been interested actively in a test that would show the ability of grade and high school pupils to recognize and correct the most common errors of the English language. The purpose of this article is to describe briefly an attempt to develop such language error tests and to show as final results three simple tests that are practically equal in value.

Determining the Errors.—A language test to be of any great value must be based squarely upon the most common mistakes made by school pupils. This has not always been done in attempts to make language error tests; partly, because data were not available until recently. Now, however, the most common errors have been charted so fully that it is no longer a matter of opinion but a matter of record.[1]

The errors selected for the present test were chosen on the basis of the language errors studies at Connersville, Indiana; Kansas City, Missouri; Boise, Idaho; Iowa Consolidated Schools; and Cincinnati, Ohio.

Determining the Method of Testing.—In determining the method of testing, it was the writer's desire to place the language errors before the child in the form in which the children themselves make the errors. It is this situation of which the child needs to become conscious. This purpose has been accomplished by putting the tests in the form of ordinary compositions, as they might be written by children. It becomes the duty of the child to recognize the errors and to correct them. The child thus becomes his own teacher. It was soon found that the test appealed to children and that they responded to it eagerly.

Some may object that it is unwise to place incorrect language forms before children; that it is false pedagogy. It is evident, however, that the tests provide a good teaching situation since improvement as a result of giving the tests is exceedingly rapid. If the tests do bring the children up rapidly, making them conscious of what is necessary

---

[1] Wilson, G. M.: Locating the Language Errors of School Children, *Elementary School Journal*, Vol. XXI, December, 1920, pp. 290–296. This is a summary of several studies.

and creating the desired interest in correct language forms, then, certainly the tests are acceptable from a pedagogical standpoint.

Getting a Good Test.—The author, working with an advanced class in Education, listed the various language errors and especially noted the 30 or 40 that were most common. As a class exercise the attempt was then made to write compositions in each of which there should be about 30 of the most common language errors. Some of these compositions were worthwhile as a basis for the formulation of tests. Finally the author was able to put three of these compositions in form for use as tests. The three compositions were entitled, "Playing Marbles," "Strawberry Time," and "A Thanksgiving Dinner." The next step was to test these tests by using them upon public school pupils. The result of this test was to demonstrate clearly that there was a wide divergence in the grade-score values from the three tests. "Playing Marbles" showed by far the most desirable distributions. The preliminary testing of the tests gave the following tentative standards:

Table I.—Tentative Standards

Grade	Test 1 "Playing Marbles"	Test 2 "Strawberry Time"	Test 3 "A Thanksgiving Dinner"
III	6	8	10
IV	10	13	17
V	14	23	19
VI	15	24	21
VII	17	24	—
VIII	18	25	—

These standards were based upon the testing of a relatively small number of pupils—410 pupils for "Playing Marbles;" 269 pupils for "Strawberry Time;" and 452 pupils for "A Thanksgiving Dinner." It was evident, however, that "Strawberry Time" was too easy for upper grade pupils. With 26 corrections possible Grade V made a median score of 23; Grade VI, 24; Grade VII, 24; and Grade VIII, 25.

In view of the fact that "Strawberry Time" and "A Thanksgiving Dinner" are not acceptable as tests because they are too easy or do not show good distributions or proper slopes from one grade to another, it will be unnecessary to reproduce them here. "Playing Marbles," however, is so acceptable from every standpoint that a copy of it will

be of interest. The form, including directions as actually used, follows herewith:

Test 1.—Correcting Language Errors. (*A game*)

Name........................	Grade........................
Town........................	School........................
Date........................	Age........................

Directions for the Game.—(To be read by the teacher, the pupils following.) This is a little *game* in which the pupil plays teacher, and corrects a composition written by a pupil. Correct by drawing a single line through words or expressions used incorrectly, and placing the correct words above them. For example if you had the following sentence to correct: "He has went home," you would correct it by drawing a single line through *went* and writing *gone* above it. Make all changes necessary to secure correctness. Work at your usual rate. You will be given reasonable time in which to complete your work. When you have finished, turn the sheet right side down and leave it on your desk. All will be permitted to finish the work unless too slow.

The composition which you are to correct follows herewith:

Playing Marbles

Marbles is a good game. I seen some boys playing the game yesterday. I went home to look for my supply of marbles. I couldn't find none, so I saw my father. I said to him: "Father, I ain't got no marbles. Will you give me a dime?" Father seen that I was in earnest, so he give me a dime. He done it willingly. Me and father is very good friends.

I started down the street. I had not went very far, when I met John Taylor. John he is a good friend of mine. He seen me leave my home, and had came to meet me. I owed him a dime, but he did not ask me to pay up. I guess he wanted me to have some marbles so as I could play with him. He had some marbles hisself.

I ask him to go to the store with me. "No," he replied, "I have got an errand to run. Can I play with you when I have did the errand?" We agreed and spent the entire afternoon together. We had lots of fun.

If the language test were to be given once only, it would be recommended that "Playing Marbles" be used as the test. The returns on it show very acceptable distributions, and good progress from one grade to another. It is a good measure of a child's ability to detect and correct errors in written language. It is so simple that it can be

used as low as Grade III; so difficult that few college students make perfect scores. Twenty-four is a perfect score.

Table II, which follows, shows a typical table of distribution for Test I, "Playing Marbles." There is the usual overlapping of ability from grade to grade but, nevertheless, good progress and no perfect scores. The returns clearly indicate a good test.

Table II.—Typical Distribution for Score of Rights
Test 1, "Playing Marbles."
(Duluth, Minn., 1918)

Score	Grades						
	III	IV	V	VI	VII	VIII	XI
0	3	3	1				
1	7	3					
2	7	2	1				
3	7	1	..	1			
4	7	6	1				
5	2	2	1				
6	3	3	3				
7	5	4	2				
8	1	5	2	5			
9	2	3	5	7	..	1	
10	5	7	6	4			
11	..	7	8	5	..	2	
12	1	3	8	11	1	2	
13	..	4	7	6	1	2	
14	..	5	11	6	..	1	2
15	..	..	13	10	2	6	2
16	..	2	10	16	2	8	
17	..	1	9	7	1	6	3
18	..	..	13	9	1	11	8
19	..	2	3	9	1	13	15
20	..	2	4	6	..	5	19
21	..	..	1	3	1	8	15
22	..	..	..	1	..	..	8
23	..	..	..	1	..	1	1
24							
Totals.......	50	65	109	107	10	66	73
Medians.....	4	10	14	15	16	18	20

Getting Three Tests of Equal Value.—It was at this stage in the development of the Language Error test that Dr. E. L. Thorndike of Teachers College was consulted. He commended the form of the test, its educational possibilities, and also its measurement possibilities. However, he asked that the research be continued until there should be three tests of equal value. This task proved a tedious one, involving a great deal of detailed statistical procedure and occasioning much delay. However, the final results gave three tests of practically equal value.

It will be of interest to note the procedure in securing these three equal tests. The first step was to evaluate all of the errors in the three previously used stories, "Playing Marbles," "Strawberry Time" and "A Thanksgiving Dinner." On the basis of tests that had been given to 1131 pupils in Sioux City, Iowa; Ambler, Pennsylvania; and Duluth, Minnesota, the error value of each error was figured for each grade. This gave for "Playing Marbles" a return as shown in Table III.

This table means that 83 per cent of the III*B* pupils failed to correct the first error in Test I, "Playing Marbles," *i.e.*, they failed to cancel "seen" in the first line and write "saw" above it. In the VI*A* grade, only 8 per cent of the pupils failed to correct this error. The errors differ greatly in difficulty, as shown by Table III. Errors 1, 3, 5, 11, 12, and 19 show gradual reduction in the higher grades. Error 17, on the other hand, is not recognized by pupils in Grade VI and below, while errors 22 and 24 are seldom recognized. It was to ascertain this difference in error value that Table III on "Playing Marbles" and similar tables on Tests 2 and 3 were constructed.

With the value of each error thus figured, it was possible to re-arrange the errors so as to distribute them equally for new stories. This re-arrangement with values for the different grades and total values is shown herewith in Tables IV, V, and VI. The initials in the following tables are explained as follows: P. M.—"Playing Marbles;" S. T.—"Strawberry Time;" T. D.—"Thanksgiving Dinner." The two errors added after the first total in each case were for the purpose of making the error values of the different stories more nearly equal and particularly to help the slope. The resulting totals are surprisingly close together and the slope from grade to grade is surprisingly uniform:

TABLE III.—PER CENT OF ERROR
Test 1. "Playing Marbles"
(Duluth)

Error	Grade								
	III*B*	III*A*	IV*B*	IV*A*	V*B*	V*A*	VI*B*	VI*A*	Total
1. *Seen* for *saw*	83	68	34	22	32	0	12	8	259
2. *None* for *any*	83	66	68	37	18	22	25	13	332
3. *Ain't* for *have*	83	62	28	11	18	11	12	8	233
4. *Got*	83	68	66	67	63	71	77	29	524
5. *No* (double neg.)	100	79	50	37	39	26	40	8	379
6. *Seen* for *saw*	100	70	65	30	46	26	23	21	381
7. *Give* for *gave*	100	95	59	44	53	39	28	25	443
8. *Done* for *did*	100	90	90	55	57	61	58	46	557
9. *Me* for *I*	100	77	59	55	53	50	33	21	448
10. *Me and father*	100	84	78	88	57	50	44	21	522
11. *Is* for *are*	100	81	59	38	46	33	19	8	384
12. *Went* for *gone*	83	77	50	32	25	33	26	8	334
13. *John, he*	100	77	97	55	58	61	65	13	526
14. *Seen* for *saw*	83	70	56	30	39	17	12	17	324
15. *Came* for *come*	83	81	66	44	53	50	42	17	436
16. *Pay up* for *pay*	100	77	81	67	61	88	44	38	556
17. *Guess*	100	100	100	100	100	100	100	100	800
18. *As* (superfluous)	100	77	87	67	61	77	56	42	562
19. *Hisself* for *himself*	100	92	62	37	32	39	28	4	394
20. *Ask* for *asked*	100	92	99	74	61	50	44	34	554
21. *Have got* for *have*	100	86	99	93	93	88	93	55	707
22. *Can* for *may*	100	92	100	100	100	94	96	76	758
23. *Did* for *done*	100	86	74	52	46	39	23	13	433
24. *Lots of*	100	95	100	93	86	100	89	55	718

TABLE IV.—STORY NO. A. BASIS FOR IN RECLASSIFIED ERRORS

Source	Value	III*B*	III*A*	IV*B*	IV*A*	V*B*	V*A* ±	VI*B* ±	VI*A* ±	VII*B* ±
S. T. 2	18	11	3	0	4	0	0	0	0	0
T. D. 23	189	59	44	15	18	17	21	9	6	
P. M. 3	233	83	62	28	11	18	11	12	8	
S. T. 4	217	66	55	34	44	4	0	14	0	0
T. D. 1	233	43	64	35	12	19	33	9	18	
T. D. 18	238	52	45	24	15	32	29	23	18	
S. T. 8	256	77	76	34	40	4	8	10	0	7
S. T. 17	269	77	59	52	68	8	0	5	0	0
S. T. 13	308	89	76	41	64	8	4	14	12	0
T. D. 7	446	76	62	52	45	63	50	50	48	
T. D. 3	370	67	69	52	24	47	44	38	29	
T. D. 21	379	59	62	46	52	49	42	25	44	
P. M. 6	381	100	70	65	30	46	26	23	21	
S. T. 20	406	89	86	57	64	19	6	33	53	0
P. M. 23	433	100	86	74	52	46	39	23	13	
T. D. 26	449	63	71	49	58	52	60	52	44	
T. D. 17	481	94	69	68	61	48	54	39	48	
S. T. 18	472	91	93	72	84	31	4	50	47	0
P. M. 13	526	100	77	97	55	58	61	65	13	
P. M. 4	524	83	68	66	67	63	71	77	29	
P. M. 18	562	100	77	82	67	61	77	56	42	
P. M. 10	522	100	84	78	88	57	50	44	21	
T. D. 22	633	85	91	79	82	81	73	68	74	
T. D. 11	669	97	80	71	82	91	94	77	77	
S. T. 11	645	82	83	92	88	81	67	55	47	50
S. T. 5	317	75	76	56	64	0	17	17	12	0
Total	10,176	2018	1788	1419	1339	1003	941	888	724	57
P. M. 7	443	100	95	59	44	53	39	28	25	
S. T. 19	227	75	62	23	44	2	4	5	12	0
Total	10,846	2193	1945	1501	1427	1058	984	921	761	57

TABLE V.—STORY NO. B. BASIS FOR IN RECLASSIFIED ERRORS

Source	Value	III*B*	III*A*	IV*B*	IV*A*	V*B*	VA ±	VI*B* ±	VI*A* ±	VII*B* ±
S. T. 22.......	164	62	41	10	24	19	5	2	0	0
S. T. 6.......	167	44	41	26	24	15	8	2	0	7
T. D. 13.......	199	37	47	28	24	16	21	16	10	
S. T. 1.......	197	57	55	23	52	0	0	10	0	0
S. T. 12.......	237	80	66	18	48	8	4	6	7	0
T. D. 10.......	250	56	58	26	33	35	17	9	16	
P. M. 1.......	259	83	68	34	22	32	0	12	8	
T. D. 9.......	312	69	67	41	36	28	29	18	24	
S. T. 26.......	321	82	79	46	56	15	0	19	24	0
P. M. 2.......	332	83	66	68	37	18	22	25	13	
S. T. 24.......	372	84	76	46	76	23	0	36	24	7
T. D. 16.......	387	63	69	45	45	48	48	30	39	
P. M. 5.......	379	100	79	50	37	39	26	40	8	
P. M. 19.......	394	100	92	62	37	32	39	28	4	
S. T. 7.......	438	88	100	57	92	0	12	36	53	0
P. M. 11.......	384	100	81	59	38	46	33	19	8	
T. D. 30.......	443	76	64	61	45	53	56	43	45	
P. M. 15.......	436	83	81	66	44	53	50	42	17	
T. D. 20.......	455	80	75	54	48	57	60	41	40	
S. T. 23.......	484	91	89	66	96	35	21	55	24	7
T. D. 25.......	506	89	82	54	48	65	50	66	52	
T. D. 12.......	548	89	62	66	70	72	73	55	61	
P. M. 20.......	554	100	92	99	74	61	50	44	34	
T. D. 15.......	573	91	96	80	61	61	63	55	66	
T. D. 29.......	691	94	85	79	76	92	89	86	90	
P. M. 21.......	707	100	86	99	93	93	88	93	55	
Total........	10,189	2081	1897	1363	1336	1016	865	888	722	21
P. M. 6.......	381	100	70	65	30	46	26	23	21	
S. T. 13.......	308	89	76	41	64	8	4	14	12	0
Total........	10,878	2270	2043	1469	1430	1070	895	925	755	21

TABLE VI.—STORY NO. C. BASIS FOR IN RECLASSIFIED ERRORS

Source	Value	III*B*	III*A*	IV*B*	IV*A*	V*B*	V*A* ±	VI*B* ±	VI*A* ±	VII*B* ±
S. T. 15.......	124	39	31	15	24	8	0	0	0	7
S. T. 3.......	169	54	24	25	36	8	4	12	6	0
S. T. 10.......	193	55	48	31	48	0	0	5	6	0
S. T. 14.......	138	39	24	67	8	0	0	0	0	0
S. T. 19.......	227	75	62	23	44	2	4	5	12	0
T. D. 2.......	268	56	69	38	27	21	35	9	13	
T. D. 14.......	269	46	64	37	27	32	31	14	18	
T. D. 5.......	324	74	58	34	33	32	40	27	26	
T. D. 6.......	325	65	58	35	48	36	33	21	29	
P. M. 12.......	334	83	77	50	32	25	33	26	8	
T. D. 28.......	341	69	65	42	36	49	33	27	20	
T. D. 19.......	370	69	62	42	45	43	42	30	37	
S. T. 25.......	423	88	86	62	96	31	16	26	18	0
S. T. 9.......	423	80	100	52	96	15	16	33	24	7
T. D. 31.......	403	81	80	59	30	47	42	30	34	
T. D. 24.......	424	89	76	49	64	48	44	23	31	
P. M. 9.......	448	100	77	59	55	53	50	33	21	
P. M. 7.......	443	100	95	59	44	53	39	28	25	
S. T. 21.......	503	88	100	59	88	26	12	48	82	0
S. T. 16.......	526	86	93	75	96	8	33	40	88	7
T. D. 4.......	537	76	75	65	73	75	67	43	63	
T. D. 27.......	522	87	78	68	61	63	65	55	45	
P. M. 16.......	556	100	77	81	67	61	88	44	38	
P. M. 8.......	557	100	90	90	55	57	61	58	46	
P. M. 24.......	718	100	95	100	93	86	100	89	55	
P. M. 22.......	758	100	92	100	100	100	94	96	76	
Total........	10,323	1999	1856	1417	1426	979	982	822	821	21
P. M. 5.......	379	100	79	50	37	39	26	40	8	
S. T. 22.......	164	62	41	10	24	19	6	2	0	0
Total........	10,866	2161	1976	1477	1487	1037	1014	864	829	21

SOME EVIDENCE OF AN ADOLESCENT INCREASE IN THE RATE OF MENTAL GROWTH

KATHERINE MURDOCK

Punahou School, Honolulu, Hawaii

AND

LOUIS R. SULLIVAN

American Museum of Natural History, New York

Many studies of mental growth have appeared very recently in our journals of psychology and education. Various phases of the subject have been attacked and defended by psychologists of note. The major issues of dispute seem to have centered about: first, the constancy of the ratio of mental development to age; that is, the constancy of the IQ, over various intervals of time and for various degrees of mental attainment; second, the comparative variability in mental development of the two sexes and of children of different life ages; and third, the limit of mental development, for normal and subnormal subjects. Such an aroused interest in the theoretical, as well as the practical, problems of mental growth will undoubtedly result in new generalizations concerning the whole topic, and the special phase of the subject with which we are here dealing will again come into prominence.

That the problem of the increase in rate of development at adolescence should for some time have been so neglected seems perhaps strange to the student of the physical development of the body. However, obvious as it is that study of adolescent changes in rate of growth is one of the primary tasks of any development study, there are good reasons why psychologists have been side-stepping that issue somewhat.

Probably one reason for the popular neglect of our subject is that mental units, sufficiently accurate for measuring such fine distinctions as our subject calls for have not yet been found. Mental development is at present usually measured in terms of mental age. That mental age steps are equal in amount is a fact not theoretically insisted upon by Terman,[1] nor practically believed in by Kelley.[2]

[1] Terman, L. M.: *Journal of Educational Psychyolog*, September, 1921. "The Binet type of scale does not necessarily presuppose equality of mental steps."

[2] Kelley, T. L.: *Journal of Educational Research*, p. 239, October, 1921. "I would say with reference to scales of the Binet type which assume equivalence of successive age intervals, I think we already have abundance of evidence to refute assumption."

In order to show changes in rate of development accurately, it is, of course, necessary that the measuring scale used *shall* be made up of equal units. Objectivity and self-equality of units, such as are illustrated by the inch and the pound, will probably never be attained in such full measure in mental measuring rods, but the earnest hope and belief of all psychologists is that mental measures may, in the future, more closely approach these ideals. Some writers have suggested that certain forms of "point scales" already excel the Binet mental age method in providing equal measuring units. Full confidence in such measures must wait upon the settlement of controversies such as have been engaged in by Freeman and Peterson, as to the proper method of equating "time" and "work." Possibly some form of the method suggested by Thurstone, the determination of standards for each mental age that state the percentage of unselected children at each life age who reach or exceed that grade of intelligence, would help to solve the equal unit problem. At any rate, it is not solved at present. This fact offers good justification to psychologists for having, in the main, avoided attempts at measuring adolescent growth accelerations. The direct use of the IQ is, of course, out of the question, since this is a measure of relative brightness, not of mental status.

Another, and probably more important, reason, why psychologists have not paid much attention to changes in rate of mental growth at adolescence, is a very practical one. The science of mental measurement is young. Matters concerning really large differences in mental ability are still unknown to many laymen. Measurers of mentality have been busy pointing these out and doing practical, necessary service. Early in the short period during which the IQ has been studied, it was conclusively shown by Terman that no *large* adolescent spurts exist. Since that was done in 1916, until perhaps within the last 12 months, little time has been spared by the psychologists studying mental differences from the practical duty of dealing with large differences in mentality, to the mere theoretical pursuit of measuring small fluctuations in the rate of mental growth.

Now the time seems to have arrived when mental growth curves should be more minutely studied. Until better measuring units are available, various devices will have to be used to show up certain features of these curves.

To study the adolescent growth curve, we have employed the simple device of making comparison between the physical and mental sex differences which occur during the development ages. Boas, in an

article on the Growth of Children, *Science*, Vol. XXXVI, has set forth clearly and concisely the laws of physical growth. The curve for rate of growth, in almost all, if not all, organs and parts of the body, he shows to have two modes. The higher mode occurs during fetal life; the second, and lesser, occurs shortly before sexual maturity is reached. Since adolescence occurs at different ages in the two sexes, the pre-adolescent increase in rate of growth occurs at different ages—that in girls occurring about 2 years earlier than the corresponding acceleration in boys. The age for minimum increase in annual growth is given by Boas as 10.3 for boys and 8.2 for girls. The age for maximum increase is 13.2 for boys and 11.2 for girls. With these facts about sex differences in physical traits in mind, it occurred to us that if a similar sex difference were found to exist in mental ability, girls exceeding boys mentally during the same ages in which they exceed them physically, that this would constitute some evidence that an adolescent increase in rate of growth is a feature of mental as well as of physical development. We already had at hand measures used by us for another study, which made this comparison between physical and mental age-sex differences very easy to make. Records of 580 boys and girls, who ranged in age from 6 to 18, were used. All of these subjects at the time the measures were made were pupils of Punahou School, Honolulu, Hawaii. This is a private school which carries the pupils from the first grade through high school. All of the 580 pupils were American or British children, of Northern European descent. Most of them had lived all their lives in the Hawaiian Islands. The measures which we had at hand were made originally by us separately for independent purposes. The anthropological data were collected by Doctor Sullivan, the mental measures by Miss Murdock. The latter measurements were made by the use of group tests. The Otis Primary Test was used for Grades I to III; the National Intelligence Tests, Forms A and B, for Grades III to IX; the Terman Group Test, for the four high school grades. Mental ages were assigned to individuals on the basis of norms furnished by the authors or publishers of the tests. Adjustment was made in order to bring ages derived from the three different tests all to the standard of the National Intelligence Tests. Ages above and below those for which norms were available, were estimated from assumptions concerning the normal curve of distribution. IQs were found for each subject by dividing the mental age by the life age. (It is true that some doubt has been thrown by Freeman and others upon the permissibility of using mental ages,

derived from group tests, to obtain IQs. For the purposes for which we have used these IQs, however, we believe there can be no objection to the way in which we have derived them.) In the case of pupils from Grades III and IX, who received mental age ratings from two tests, an average of the two was used. The measures for weight were expressed in terms of pounds; those for stature, in terms of centimeters.

The results of the comparison of physical and mental measures for the two sexes at successive ages, are given in the accompanying tables. The average weight, stature, and intelligence quotient for boys and for girls at each age from 6 to 18 are given, and also the amount in each age group by which the girls exceed the boys, or the boys the girls, in weight, stature, and IQ. In the case of the IQ, the excess of girls over boys, and of boys over girls, is given for each age separately, and also as smoothed averages, in which the average given for each age group is in reality the average for the boys (or the girls) of that age group in combination with the age just younger and the one just older. Comparison of the excess columns is very striking in revealing that there is a similarity between the mental and physical ages of development. If we confine our attention to the "smoothed" averages for the IQs, and compare these with the physical measures, we find that the direction of excess, of boys over girls, or vice versa, at different ages, is as constant for mental and physical measures as it is between the two physical measures themselves. From 8 years of age until 13, the girls excel, physically and mentally. Thereafter they are behind the boys. The rough, or unsmoothed, excesses tell about the same story. The greatest exception occurs at the age of 18, where girls are seen to excel the boys, mentally, by an average of 2.4 IQ. It occurs to us as possible that the brighter of the 18-year-old boys may be sent to college more often than the bright girls of 18—parents disliking to have their daughters go so far from home at this early age. However this is only a supposition. The smallness of the groups naturally would result in irregular results for differences which are so small as those which we are attempting to measure. We would not, in fact, feel justified in presenting our data at all, based, as they are upon such limited numbers of cases, were it not for the fact that so many age groups unite in confirming the evidence.

Our own interpretation of our results is that they furnish evidence that a pre-adolescent increase in rate of mental growth occurs at the same time in the development of each sex that the physical increase in

development occurs. Whether or not boys excel girls mentally, as they do physically, after, and to a slight extent, before the adolescent growth periods, is a question on which our data hardly shed enough light for us to form an opinion. Neither do we feel justified in assuming anything about the amount of the adolescent mental growth acceleration, except that it is probably much smaller, comparatively, than the physical "spurt." The unit of mental age scales, by its very definition, is of such a nature that it tends to conceal any differences in rate of mental growth. Eleven years mental age means the mental age of the average 11-year-old child. If, on the average, children should develop little mentally from 10 to 11 years of age, and develop much from 11 to 12, properly arranged mental scales, of the age standard type, would entirely conceal such change in rate of development. (The IQ as a measure, is similarly limited, with the additional restriction, when it is used for purposes of studying mental development, that only subjects of the same life age can be compared. In our present study, we have used IQs, instead of mental age, only because our measures already had been converted into this form for other purposes. Since our comparisons all are between groups composed of individuals, who are of the same life age, the results are identical, whether mental age or IQs are used.) Since then changes in mental growth are concealed, rather than shown up, by the use of mental scales, it is impossible for us to arrive at a decision as to the amount of adolescent growth acceleration. Its existence at all, by the use of such scales, could not have been determined, were it not for the sex differences which we found. Mental scales have been standardized by the use of results obtained from the scores of both boys and girls.

On the whole our results seem to be in harmony with those of other investigators. Porteus, in his maze studies, found some correlation between physical and mental development, with a consequent superiority, in mental ability, of girls to boys from the age of 11½ to 13 years, and inferiority at both earlier and later ages. Yerkes and Bridges, by the use of their Point Scale for Measuring Intelligence, found boys superior to girls from 8 to 11, girls superior at 12 and boys again superior from 13 to 15. That the ages for the girls' superiority in our study come somewhat earlier than in these others, is possibly due to the fact of the early physical development, brought about by the warm climate or the social status of our subjects. Terman's findings seem also to yield substantially the same results as ours On p. 75 of "The Stanford Revision of the Binet-Simon Scale" he says:

	Averages								Sex differences in averages							
	Number of cases		Weight in pounds		Stature in centimeters		IQ		Excess in weight		Excess in stature		Excess in IQ		Excess in IQ (smoothed average)	
Age	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Girls exceed boys	Boys exceed girls	Girls exceed boys	Boys exceed girls	Girls exceed boys	Boys exceed girls	Girls exceed boys	Boys exceed girls
6	8	12	50	48	121	119	109.1	104.9	..	2	..	2	....	4.2	...	
7	22	18	54	51	125	122	112.7	107.7	..	3	..	3	....	5.0	...	2.2
8	18	25	58	59	130	129	99.7	103.2	1	..	..	1	3.5	...	2.0	
9	29	23	63	68	133	136	102.8	110.1	5	..	3	..	7.3	...	6.3	
10	32	22	68	75	136	142	105.0	115.6	7	..	6	..	10.6	...	5.7	
11	33	17	81	89	145	149	113.1	113.1	8	..	4	..	0.0	...	4.9	
12	26	28	87	92	151	152	109.0	113.2	5	..	1	..	4.2	...	1.2	
13	27	28	97	111	155	160	113.2	113.0	14	..	5	..	...	0.2	1.0	
14	34	35	112	113	164	161	110.0	109.6	1	..	..	3	....	0.4	...	2.0
15	24	19	129	118	171	164	110.9	106.0	..	11	..	7	....	4.9	...	1.7
16	18	32	135	123	173	164	107.9	108.7	..	12	..	9	0.8	...	...	2.6
17	14	18	135	122	175	165	114.4	108.5	..	13	..	10	....	2.9	...	0.1
18	7	13	140	122	175	165	109.9	112.3	..	18	..	10	2.4			

> "In the main, therefore, the school progress of our subjects agrees with the intelligence tests, with the teachers' estimates of intelligence, and with the teachers' judgments of the quality of the school work, in showing a sex difference which is in favor of the girls before 14, and in favor of the boys thereafter."

This statement, however, is immediately followed by Terman with an explanation of what he thinks is the probable reason for his findings. He believes the apparent superiority of boys over 13, to girls, to be due to the effect of selection of his subjects. All of these were pupils in the elementary school, and his belief is that more girls than boys, of 14 years of age, had been advanced to high school. His final conclusion therefore is different from ours. It is that "the only possibility seems to be that the apparent superiority of boys at the age of 14, as well as also their diminished inferiority at 13, is due solely to the uneven selection which has taken place at these ages." However this may be in the case of Terman's subjects, it certainly is not true for ours that the superiority of the boys after 13 years of age is due to a selective influence, which places more girls in high school, for in our study high school students as well as those in elementary school are tested. All pupils of the ages 6 to 18 in the whole school, which consists of 12 grades, were included in our study (except those of other races). Another important study, whose results harmonize with our conclusions, is Mrs. Pressey's study of sex differences, in which she finds girls slightly superior to boys in mental ability. Mrs. Pressey's pupils were elementary school pupils, therefore mostly below 14 years of age.

We hope that in the future workers who compare the mental ability of the two sexes, will present their results in such a way that the sexes can be compared age for age during the developmental years. Further studies of this sort, including measures of many more subjects, alone can corroborate, or refute, our tentative conclusion that for both sexes there is an increased rate of mental as well as physical development for several years prior to the attainment of sexual maturity.

TENTATIVE ORDER OF DIFFICULTY OF THE TERMAN VOCABULARY WITH VERY YOUNG CHILDREN

MARGARET V. COBB

Institute of Educational Research, Teachers College, Columbia University

In the course of a study of group intelligence tests in Grade I, which involved Binet examining, it became evident that the words in the Vocabulary Test (Stanford Revision) show an order of difficulty for small children which is noticeably different from that indicated on the blank. Tabulation of the scores on the separate words gave the following information, which may prove useful to others who are examining kindergarten and Grade I children.

The examining was done by five persons, including myself, the others being Miss Helen Davis, Director of Measurements and Special Education in the public schools of Jackson, Michigan, through whose courtesy the records were made available, and three of the kindergarten teachers of Jackson whom she had trained for the work and who had shown special aptitude for it. The children were examined in the kindergarten during the first semester of the school year 1920–1921, and in Grade I during the second semester of the same year. With a few children the test was begun with word 6 in each list; the previous words were omitted but given credit, if there were no failures before word 11 ("Roar" and "haste" were, however, always given.). The test was stopped when the child missed five successive words in each list. Three or four of the scoring conventions observed in addition to the rules in "The Measurement of Intelligence" should probably be noted here, since a different rule might have altered the result:

Bonfire.—Full credit, any definition which gave a hint of distinction between a bonfire and other fire, as "A fire outdoors," "A big fire," "Go burn things up," etc. Half credit, definitions of fire alone, as "Furnace fire," "Fire," "In a stove," etc.

Haste.—Full credit for definition involving hurry or speed. Half credit for quotation from school song ("haste away," etc.) or definition as "go," "fly," etc.

Afloat.—Full cedit for definition involving floating on surface. Half credit for definition as moving along, swimming, carried along, etc.

Eyelash.—Half credit for "hair over eye" if child points to eyebrow instead of lashes.

The distributions of total scores on the vocabulary test, and of chronological and mental ages at the time of examination, are shown

in Tables I, II and III. Table IV gives the data concerning each word of the vocabulary, including the number of times each was given and the number of successes, of failures, and of half credits in the kindergarten and the lower and upper Grade I. With the present word order, the following words, unknown to any of these children, were needlessly asked:

First Column	Times Asked	Second Column	Times Asked
16. skill	35	19. forfeit	7
17. ramble	18	20. sportive	5
18. civil	14	23. shrewd	1
21. juggler	7	24. repose	1
22. regard	4	25. peculiarity	1
23. stave	1		
24. brunette	1		
25. hysterics	1		

On the basis of these figures, it is suggested that when the vocabulary test is given to young children the following word-order will be found useful, and will give greater certainty in making inferences as to when it is safe to omit the easiest words, and when the child has been carried far enough down the list. It also makes the two lists more nearly equal in difficulty, so that the use of one alone is somewhat more accurate. (It is my belief that this procedure is, however, almost never advisable.) Mimeographed sheets may easily be made up in this order, or in whatever similar order anyone wishes to derive from the data.

	Credits		Credits
1. straw	104	1. orange	106
2. envelope	103	2. bonfire	97
3. gown	83	3. puddle	93
4. tap	87½	4. rule	81
5. scorch	73½	5. roar	73
6. eyelash	66	6. pork	55
7. afloat	39	7. health	45
8. impolite	32	8. plumbing	32
9. copper	23	9. haste	23
10. nerve	14	10. guitar	21
11. dungeon	9½	11. muzzle	13
12. curse	9	12. misuse	7
13. southern	7	13. snip	6
14. lecture	4	14. reception	4½
15. mellow	4	15. noticeable	3½
16. insure	3	16. quake	2½
17. outward	2½	17. treasury	2½
18. apish	2	18. crunch	I
19. ramble	0	19. majesty	

TABLE I.—DISTRIBUTION OF TOTAL SCORES, VOCABULARY

Total scores	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Kindergarten	1	..	3	..	4	6	..	..	1	1	1	..	1	..	..	..	1	..	..	..
IB	..	..	3	..	3	3	2	6	5	4	2	3	1	2	..	..	1	..	1	1
IA	..	1	..	..	5	2	4	7	7	1	5	3	5	3	3	..	1	1	2	..
Total	1	1	6	0	12	11	6	13	13	6	8	6	7	5	3	0	3	1	3	1

TABLE II.—DISTRIBUTION OF MENTAL AGES AT TIME OF EXAMINATION

Mental ages	5^8	5^{10}	6^0	6^2	6^4	6^6	6^8	6^{10}	7^0	7^2	7^4	7^6	7^8	7^{10}	8^0	8^2	8^4	8^6	8^8	8^{10}	9^0	9^2
Kindergarten	2	1	4	2	3	1	4	1	1													
IB	..	..	1	1	1	2	1	3	4	4	4	3	4	2	3	2	0	2				
IA	..	..	..	..	1	3	3	4	5	4	10	8	2	0	4	1	3	1	0	0	0	1
Total	2	1	5	3	5	6	8	8	10	8	14	11	6	2	7	3	3	3	0	0	0	1

TABLE III.—DISTRIBUTION OF CHRONOLOGICAL AGES AT TIME OF EXAMINATION

Chronological ages	5^2	5^4	5^6	5^8	5^{10}	6^0	6^2	6^4	6^6	6^8	6^{10}	7^0	7^2	7^4	7^6	7^8	7^{10}	8^0	8^2	8^4	8^6	8^8	8^{10}	9^0
Kindergarten	1	3	1	4	4	3	2	0	1															
IB	..	..	..	..	1	5	1	3	9	1	9	2	2	2	0	1	0	0	0	0	1			
IA	..	..	..	..	..	..	..	..	1	8	11	7	6	3	6	2	3	1	0	1	0	0	0	1
Total	1	3	1	4	5	8	3	3	11	9	20	9	8	5	6	3	3	1	0	1	1	0	0	1

TABLE IV

First Column

Word order on blank	Full credit			Half credit			No credit			Times asked			Total credits	Total times asked
	Kindergarten	IB	IA	Kindergarten	IB	IA	Kindergarten	IB	IA	Kindergarten	IB	IA		
1. gown	14	28	41	..	..	..	5	9	9	19	37	50	83	106
2. tap	12	29	46	..	1	..	7	7	4	19	37	50	87½	106
3. scorch	12	25	35	1	..	2	6	12	13	19	37	50	73½	106
4. puddle	18	32	43	..	..	..	1	5	7	19	37	50	93	106
5. envelope	18	36	49	..	..	..	1	1	1	19	37	50	103	106
6. rule	13	26	42	..	..	..	6	11	8	19	37	50	81	106
7. health	7	16	22	..	..	..	12	21	28	19	37	50	45	106
8. eye-lash	8	24	26	1	6	9	10	7	15	19	37	50	66	106
9. copper	4	8	8	2	1	3	10	28	39	16	37	50	23	103
10. curse	1	4	4	..	..	..	10	33	46	11	37	50	9	98
11. pork	0	10	20	1	22	27	0	5	3	1	37	50	55	88
12. outward	0	1	0	..	2	1	..	25	37	..	28	38	2½	66
13. southern	..	0	5	..	2	2	..	26	31	..	28	38	7	66
14. lecture	..	1	3	..	..	..	..	25	35	..	26	38	4	64
15. dungeon	..	3	6	..	1	..	..	22	33	..	26	39	9½	65
16. skill	..	0	0	..	..	..	..	16	19	..	16	19	0	35
17. ramble	..	0	0	..	..	..	..	7	11	..	7	11	0	18
18. civil	..	0	0	..	..	..	..	5	9	..	5	9	0	14
19. insure	..	1	1	..	1	1	..	12	23	..	14	25	3	39
20. nerve	..	0	3	..	9	13	..	5	11	..	14	27	14	41
21. juggler	..	0	0	..	..	..	..	1	6	..	1	6	0	7
22. regard	..	0	0	..	..	..	11	1	3	..	1	3	0	4

Second Column

Word order on blank	Full credit			Half credit			No credit			Times asked			Total credits	Total times asked
	Kinder-garten	IB	IA	Kinder-garten	IB	IA	Kinder-garten	IB	IA	Kinder-garten	IB	IA		
1. orange	19	37	50	..	..	..	..	..	..	19	37	50	106	106
2. bonfire	10	36	45	9	..	3	..	1	2	19	37	50	97	106
3. straw	18	37	49	..	..	..	1	..	1	19	37	50	104	106
4. roar	8	27	37	..	..	..	11	10	13	19	37	50	72	106
5. haste	1	1	15	1	1	10	17	35	25	19	37	50	23	106
6. afloat	2	6	15	7	11	14	10	20	21	19	37	50	39	106
7. guitar	4	5	12	..	..	..	15	32	38	19	37	50	21	106
8. mellow	2	1	1	..	..	..	14	36	48	16	37	49	4	102
9. impolite	3	13	15	..	2	..	11	21	34	14	36	49	32	99
10. plumbing	4	12	12	1	1	6	8	22	32	13	35	50	32	98
11. noticeable	0	1	1	..	1	2	1	22	36	1	24	39	3½	64
12. muzzle	0	4	9	..	..	..	1	22	36	1	26	45	13	72
13. quake	..	0	0	..	4	1	..	10	21	..	14	22	2½	36
14. reception	..	2	2	..	0	1	..	10	21	..	12	24	4½	36
15. majesty	..	0	0	..	0	1	..	9	19	..	9	20	½	29
16. treasury	..	0	1	..	1	2	..	7	16	..	8	19	2½	27
17. misuse	..	0	7	..	..	0	..	10	13	..	10	20	7	30
18. crunch	..	0	1	..	..	..	..	6	13	..	6	14	1	20
19. forfeit	..	0	0	..	..	..	..	1	6	..	1	6	0	7
20. sportive	..	0	0	..	..	..	..	1	4	..	1	4	0	5
21. apish	..	2	0	..	..	..	..	1	11	..	3	11	2	14
22. snip	..	4	2	..	..	..	..	3	15	..	7	17	6	24

The larger changes in order seem to be due to the fact that certain more abstract terms, such as *afloat*, *haste*, *mellow*, *noticeable*, *outward* and *skill*, are relatively more difficult at these early ages; and that some of the more concrete terms, such as *dungeon*, *envelope*, *eyelash*, *nerve*, *plumbing*, *pork* and *snip*, are learned earlier and are (relative to an older group having an equally small vocabulary) more likely to be known. Smaller changes may be due to the fact that the number of cases tabulated is small, though in general the lists from the subgroups confirm one another. The first four words appear in the same order with the kindergarten and I*B* grouped together as with the I*A* children; the first 13 words for the two groups are the same words, though the order is different. However, many tabulations are required before the best order for any limited group can be attained. It is suggested that others who have accumulated Stanford-Binet records for children of this age might tabulate and publish their data in such form that eventually they could be combined, and the best order finally determined.

SOME RETESTS WITH THE STANFORD-BINET SCALE

KATE GORDON

Bureau of Children's Aid, California Department of Finance

In a recent issue of the *Journal of Educational Psychology*, (September, 1921), an important summary was made of several studies on the constancy of the Stanford-Binet IQ as shown by retests. The number of cases reported in the present paper is small, but each additional bit of evidence on so vital a question seems to be worth having. Moreover, my results, as far as they go, indicate a difference in the results of retests on the two sexes.

All of the original tests on the group here reported, and all of the retests were made by me. The children tested are all of one race—the Hebrew—and all are living in the same environment, an orphanage for Jewish children. The original tests were made in October, 1918, when a complete mental survey of the institution was in progress. In December, 1919, those were examined who had been admitted since the previous date. In August, 1921, another complete survey was made of the same institution. The results given below include all of the children who were tested twice in the course of this procedure, namely, 44 persons. For 34 children the interval between tests was approximately 2 years 10 months; for 9 children about 1 year 8 months, and for 1 child, it was 1 year 3 months.

The number being small, I give the individual records in Table I.

In the cases of Boy No. 23, and Boy No. 27 I believe that the loss in IQ is due to the fact that these boys, who passed most of the 18-year-old tests of the scale, had no chance to earn a higher IQ because of the upper limit of the scale. On the other hand, Boy No. 17 also passed most of the 18-year-old tests, and had he been able to earn more, his gain in IQ would have been still greater.

The total distribution of changes in IQ is given in Table II. Their average is 6.8.

In spite of the occasional large changes indicated in Table II there is on the whole a very substantial agreement between the results of the first and second tests of this group, the correlation being r-0.84, as shown in Table III.

A glance at Table I reveals the fact that most of the losses were with the girls and most of the gains with the boys. I can assign no

TABLE I

Boy number	Age at Test I		Intelligence quotient on Test I	Age at Test II		Intelligence quotient on Test II	Points lost	Points gained
	Years	Months		Years	Months			
1	4	2	100	7	0	95	5	
2	5	0	110	7	11	112	..	2
3	5	8	100	7	4	91	9	
4	6	4	95	8	0	94	1	
5	7	3	89	8	11	93	..	4
6	8	5	85	10	1	98	..	13
7	8	5	72	11	2	69	3	
8	8	7	95	11	5	99	..	4
9	8	10	89	11	8	96	..	7
10	9	0	92	10	8	103	..	11
11	9	1	122	11	10	138	..	16
12	9	1	119	11	10	118	1	
13	9	9	95	12	7	87	8	
14	9	10	92	11	6	93	..	1
15	9	10	94	12	8	91	3	
16	10	4	100	13	2	99	1	
17	10	7	125	13	5	141+	..	16
18	10	7	94	13	5	106	..	12
19	11	5	100	14	3	110	..	10
20	11	5	91	14	2	105	..	14
21	11	6	107	14	4	108	..	1
22	11	7	71	12	10	77	..	6
23	11	10	130	14	7	122+	8	
24	12	0	79	14	10	83	..	4
25	12	7	98	15	5	111	..	13
26	12	7	98	15	5	101	..	3
27	12	11	126	15	10	117+	9	
28	13	0	92	14	8	99	..	7
Girl number	**Age at Test I**		**Intelligence quotient on Test I**	**Age at Test II**		**Intelligence quotient on Test II**	**Points lost**	**Points gained**
	Years	**Months**		**Years**	**Months**			
1	5	4	103	8	2	100	3	
2	6	4	76	7	11	80	..	4
3	6	4	79	7	11	78	1	
4	8	2	114	10	11	110	4	
5	8	3	109	11	0	106	3	
6	8	5	109	11	3	104	5	
7	8	11	121	11	9	118	3	
8	8	11	95	11	9	81	14	
9	10	7	83	13	5	75	8	
10	11	6	103	14	4	91	12	
11	11	8	89	14	6	80	9	
12	11	10	92	14	8	89	3	
13	12	10	101	15	8	98	3	
14	12	10	94	15	8	84	10	
15	13	2	116	16	0	93	23	
16	13	7	97	16	5	103	..	6

reason for this unless it be true that the girls come to an earlier stop in mental growth. Girl No. 15, who has a drop of 23 points seems an entirely normal child. Naturally one does not regard the matter as proved by 16 cases, but these are sufficiently arresting to suggest that it might be well for the records of boys and girls to be listed separately in reporting on the subject of retests.

"THE CONSTANCY OF THE IQ" AGAIN

FLORENCE M. TEAGARTEN

University of Pittsburg

All persons concerned with individual mental testing have been following with keen interest, in recent months, the discussion running through psychological literature relative to the constancy and diagnostic value of the IQ. Considerable evidence has been brought forth which claims to prove that the IQ can not be depended upon to give consistent results in subsequent tests. It appears, however, that in many of the investigations thus reported there have been present very conspicuous factors which have not only militated against the constancy of the IQ but also against the validity of any argument apparently inherent in the research.

One of these factors is the use of different tests or of different revisions of the same test. No very convincing scheme has been offered as yet for transmuting test results obtained by one revision or one test into supposedly comparable results by another revision or another test.

Again, some recent investigations conducted upon the feeble-minded claim to have proven that no dependence can be placed in the predictive value of the IQ. It is not the purpose of this report to dispute this point in the case of *institutional feeble-minded and of all ages.* For such cases the IQ may or may not be of significance but whether it be or not, are we justified in assuming that similar results are bound to accrue from tests of a school population; *e.g.*, since, it appears, we can not come to any agreement upon such fundamental questions as the nature of intelligence or the probable duration of its development it does seem a little beside the point to make sweeping assertions to the effect that laws of mental development of institutional feeble-minded as determined by their IQs are therefore the laws of mental development for the population at large.

The point which the writer here wishes to make is that what we probably need most right now is an *intelligent evaluation* of individual IQs before we can make intelligent comparisons of IQs. Probably sufficient evidence could be produced to show that in many cases where the bare IQ, stripped of all interpretation and standing in its mathematical nakedness, has appeared to change upon subsequent tests the very same IQ, clothed with a little common sense, could easily be identified later on. Two cases of this nature are here cited.

In November, 1921, two students taking training under the writer in administering Binet Tests (Standard Revision) examined, among others, two children in a certain practice school. The students had had several weeks of training. While the writer feels that there is always very grave danger in making serious comparisons of tests administered by different people, yet the dangers in this case were as small as it is ever possible for them to be since the students had their training from the one who made the retests and since the students' work was very carefully checked over with them on points of scoring, etc., following the examination. In March, 1922, for certain administrative reasons the writer was asked to go into the practice school and retest these two children, Jack and Mary.

The second examination showed that in the interval between November and March Jack's IQ had risen from 83 to 90+ and Mary's from 119+ to 128+. In the case of Jack the figures seemed to show that in something over 3 months he had risen from being a "dull normal" to a "normal" child and in exactly 4 months Mary had changed from a "superior" to a "very superior" child. Knowing the technique of the student testers, the writer did not feel warranted in passing off the matter by saying: "Oh, well, it is just a bit of the personal equation showing up." On the other hand, having a firm conviction in a reasonable constancy of the IQ she did not feel that the case could be dismissed without seeking some reason for this apparent lack of conformity.

In order that the reader may make the analysis for himself a brief chart of both tests for Jack and for Mary are here given. On the November test, Jack was chronologically 9-10 and mentally 8-2. On the March test his chronological age was 10-2 and his mental age 9-3, or a gain of 13 months. Mary, on the November test, was chronologically 7-7 and mentally 9-1. On the March test she was chronologically 7-10 and mentally 10-1, or a gain or exactly one year. The chart shows in the case of each part of the test the credit given (plus or minus) for that part together with the score for each portion of it. For example, in Year IX Test 2 (arrangement of weights) three trials are allowed. A plus score requires that two out of the three be correct. In the case of Jack's first examination —(− − −) means that he did not score on that test and that he failed all three trials. In his retest his score for the same test is +(+ − +) which means that his score was plus because he had correct two attempts out of three.

From a comparison of these reports for Jack it is evident that his

JACK

Test Nov. 22, 1921.
Chronological age 9-10: Mental age 8-2: Intelligence quotient 83.

Year VII

Tests all passed.

Year VIII

1. Ball and field........ −
2. 20-0........ +
3. Comprehension........ − (− + −)
4. Similarities two things........ − (? − − −)
5. Definitions........ + (+ + + +)
6. Vocabulary........ + (26)

Year IX

1. Date........ − (+ + + −)
2. Weights........ − (− − −)
3. Makes change........ + (+ + −)
4. Four digits backwards........ − (− − −)
5. Three words in sentence........ + (+ + −)
6. Rhymes........ + (+ + +)

Year X

1. Vocabulary........ − (26)
2. Absurdities........ − (+ − − + +)
3. Designs........ +
4. Reading and report........ − (14 memories, 30″, 3 mistakes)
5. Comprehension........ − (+ − −)
6. Sixty words in three minutes........ − (55)

Years XI, XII

1. Vocabulary........ − (26)
2. Abstract words........ − (+ − − − ?)
3. Ball and field........ −
4. Dissected sentences........ − (− − −)
5. Fables........ (Omitted)
6. Five digits backwards........ − (See IX, 4)
7. Picture interpretation........ − (? − − omitted)
8. Similarities 3 things........ − (See VIII 4)

Years XIII, XIV

No tests given since all of XI, XII failed.

Test Mar. 6, 1922.
Chronological age 10-2: Mental age 9-3: Intelligence quotient 90+.

Year VII

Tests all passed.

Year VIII

1. Ball and field........ + Gain 2 months
2. 20-0........ +
3. Comprehension........ + (+ + +) Gain 2 months
4. Similarities........ − (+ − − −)
5. Definitions........ + (+ + + +)
6. Vocabulary........ + (27)

Year IX

1. Date........ − (+ + − +)
2. Weights........ + (+ − +) Gain 2 months
3. Makes change........ − (+ − −) Loss 2 months
4. Four digits backwards........ + (+ − +) Gain 2 months
5. Three words in sentence........ + (+ + omitted)
6. Rhymes........ + (+ − +)

Year X

1. Vocabulary........ − (27)
2. Absurdities........ + (+ + + + omitted) Gain 2 months
3. Designs........ +
4. Reading and report........ + (14 memories, 25″, 0 mistakes) Gain 2 months
5. Comprehension........ − (− − −)
6. Sixty words in three minutes........ − (47)

Years XI, XII

1. Vocabulary........ − (27)
2. Abstract words........ − (+ − − − ?)
3. Ball and field........ −
4. Dissected sentences)........ − (− − −)
5. Fables........ − (+ − − − −)
6. Five digits backwards........ − (− − −)
7. Picture interpretation........ + (+ + + +) Gain 3 months
8. Similarities three things........ − (− − − omitted)

Total gain 13 months

Year XIII, XIV

Abbreviated test (4 tests) given—all failed.

MARY

Test Nov. 7, 1921.

Chronological age 7-7. Mental age 9-1. Intelligence quotient 119+.

Year VIII

Tests all passed.

Year IX

1. Date .... − (+++−)
2. Weights .... − (−+−)
3. Makes change .... + (++−)
4. Four digits backwards .... + (−++)
5. Three words in sentence .... − (+−−)
6. Rhymes .... + (+++)

Year X

1. Vocabulary .... − (24)
2. Absurdities .... + (+++++)
3. Designs .... +
4. Reading and report .... − (14 memories, 40″, 0 mistakes)
5. Comprehension .... − (−+−)
6. Sixty words in three minutes .... − (55)

Years XI, XII

1. Vocabulary .... − (24)
2. Abstract words .... − (−−−−−)
3. Ball and field .... −
4. Dissected sentences .... − (−− omitted)
5. Fables .... − (−−−−−)
6. Five digits backwards .... − (−−−)
7. Picture interpretation .... − (+−−−)
8. Similarities three things .... + (−++−+)

Years XIII, XIV

No tests given since all about 1 failed in Years XI, XII and nature of tests at this level indicated failure.

Test Mar. 7, 1922.

Chronological age 7-10. Mental age 10-1. Intelligence quotient 128+.

Year VIII

Tests all passed.

Year IX

1. Date .... + (++++) Gain 2 months
2. Weights .... + (−++) Gain 2 months
3. Makes change .... + (++−)
4. Four digits backwards .... + (+−−)
5. Three words in sentence .... + (++−) Gain 2 months
6. Rhymes .... + (+++)

Year X

1. Vocabulary .... − (22)
2. Absurdities .... + (++−++)
3. Designs .... +
4. Reading and report .... + (10 memories, 23″, 0 mistakes) Gain 2 months
5. Comprehension .... + (++−) Gain 2 months
6. Sixty words in three minutes .... + (60) Gain 2 months

Years XI, XII

1. Vocabulary .... − (22)
2. Abstract words .... − (+−−−−)
3. Ball and field .... + Gain 3 months
4. Dissected sentences .... − (−+−)
5. Fables .... − (−−−−−)
6. Five digits backwards .... − (−−−)
7. Picture interpretation .... − (−−− omitted)
8. Similarities three things .... − (−+−−+) Loss 3 months

Total gain 12 months

Years XIII, XIV

Abbreviated test (4 tests) given—all failed.

actual mental gain is much more nearly negligible than his 13 months accelerated credit would at first glance seem to indicate. In fact there are actually only two tests that were *entirely* failed the first time and *entirely* passed the second time, namely VIII-1, Ball and Field (counting 2 months credit) and XII-7 Picture Interpretation (counting 3 months credit). Five months credit then were undoubtedly gained in the interval of a little over 3 months. If Jack's intelligence were developing at anything like a normal rate we could probably admit a gain of 5 months in more than 3 months "and no questions asked." But where did his other 8 months credit come from? Ah! Here comes the need for analysis and interpretation, for some of them came from tests which he barely failed the first time (+ − −) and barely passed the second time (+ + −) and from similar "marginal" gains. For example, on the second test he received 2 months additional credit in the VIII Year Comprehension Test for a record of (+ + +) as against his previous record of (− + −). In fact, if his record on the second test had been only (+ + −) he would have received the additional 2 months credit since passing requires only 2 out of 3 successes at this point. Since, however, he had been able to pass even one of these tests in the first examination we can not say that he showed any unusual development to be able to pass two more 3 months later. Or again, consider Test IX-2 (Weights) where his first record was (− − −) and his second was (+ − +), and yet he gets 2 months additional credit for this partial ability which counts as a complete success. Exactly the same thing is true in Test IX-4. In Test X-2, the first record of (+ − − + +) would have been plus if there had been *one* more success which he did get 3 months later. In other words, he receives 2 months credit for detecting *one* more absurdity on the retest than he had on the previous test. In the Reading and Report Test Year X-4, success requires that the selection be read in 35 seconds allowing 2 mistakes in reading and that 8 "memories" be reported. On Jack's first test his record was 14 memories, 30 seconds, and 3 mistakes. The 3 mistakes in reading were fatal or at least the third one was, for he is allowed only 2. Now, if he can manage 3 months later to make only 2 mistakes in reading his mental age will go up 2 months. Does he do it? Fourteen memories, 25 seconds and no mistakes in reading. And for this improvement, after 3 months of school training, he gets 2 months credit toward his mental age.

On the other hand note the tests where on the re-examination he

either did more poorly than he did on his first test or else made mistakes different from the ones made before. Note IX-1, IX-3, X-5, X-6.

The 9-months credit earned on VIII-1, IX-2, IX-4, XII-7 we do not begrudge Jack, for they stand for either partial or complete success on the second test as against *absolute failure* on the first. Tests VIII-3, X-2, X-4, however, represent rather questionable gain since they are earned through such slight increments over partial successes.

For the sole purpose of an interesting if entirely unwarranted and unscientific experiment, let us see what happens to Jack's IQ if we add these last named and questionable 6 months credit to his Mental Age on the first examination, as though they had been successes instead of only partial successes. This would give an IQ of 88 on the first examination as against an IQ of 90 on the second. (As a matter of fact, this is probably really nearer the truth than the unaltered data unless we put intelligent interpretation on these data.) Would this much difference in IQs seem to indicate that the IQ is worthless or would it hint that Jack is a little below "average" but near the upper end of the "dulls?" This seems to be supported by several facts, namely that he has the vocabulary of only an 8-year-old and that in 3 months he increased that vocabulary score by only 1; by the fact that his basal year was as low as 7 considering that he is a 10-year-old child; and by the character of some of his responses.

And now to turn to the case of Mary. We see that she gained 15 months and lost 3 on her re-test, making an apparent gain of 12 months in exactly 4 months to the day. By inspection of her chart, however, we find that the XII year Ball and Field success on the second test is absolutely the only plus score where there had been an *entire* failure on the first test. In fact, the marginal failures in the first test are very conspicuous when compared with the corresponding successes on the second test: IX-1, IX-2, IX-5, X-4, X-5, and X-6. Also as in the case of Jack, we find here some partial failures on the second test where there had been none before and also one final minus score as against a plus (XII-8).

Now that we have Mary's gain accounted for, what is to be said about her *real* IQ? Is it nearer 128 or 119? If an IQ of 120 is to be considered a dividing line (Terman 110–120 "superior;" 120–140 "very superior") to which group does the child really belong? With an IQ on the first test only one point below the lower limits of "very superior" and containing 5 tests totalling 10 months credit in each of which one single more success in addition to the ones already earned

would have been rewarded with a "plus" score, one sees the necessity for a critical evaluation of this 119 IQ. Certainly this child is far superior to a child with an IQ of 119 or even 120 who fails *entirely* on such tests as IX-1, X-4. (Note Mary's minuses on these tests, composed of some plus and some minus scores.) Furthermore there was every indication throughout the test administered by the writer of Mary's great superiority as, for example, her self-criticism, choice of words and the like. While the vocabulary is not high and the child does not talk a great deal (as does Jack) yet she "delivers the goods." She is undoubtedly a "*very superior*" child despite the IQ of 119. Fortunately in this case we have one other line of evidence upon which to fall back and that is a test made by still another examiner about a year before the first of our two tests. Without comparing procedure, scoring, etc., it is interesting just to note in passing that the IQ on that test was 125—"very superior."

In conclusion the writer wishes to state as an opinion formed after a considerable testing experience (although the data cited above do not present all the evidence) that:

1. In order to be of any value whatever every IQ must be critically evaluated for the purpose of seeing what factors have contributed to it or militated against it and of seeing what "marginal" failures and successes there may be.

2. Slight variations in IQs from different testings are to be expected and may mean very little. In every case comparisons of IQs should be made in the light of (1) above.

3. In the main the IQ of children up to about 16 years of age (probably excluding the lowest grades of defectives who soon find themselves in institutions) is reasonably constant and reliable and is highly valuable for diagnostic purposes.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY CECILE COLLOTON

Department of Educational Psychology, The Lincoln School of Teachers College

INTELLIGENCE TESTS

Intelligence as Related to Nationality. Gilbert L. Brown. Journal of Educational Research, 1922, April, 324–327. Nine hundred and thirteen children of foreign parentage tested by the Stanford-Binet show wide range of intelligence Germanic groups—Norwegian, German, Swede, English and Austrian test higher than any of the non-Germanic groups.

The Relative Progress of VII-B Groups Sectioned on the Basis of Ability. W. W. Theisen. Journal of Educational Research, 1922, April, 295–305. Advantages of grouping pupils on basis of ability. Needed changes in curricula, achievement standards, and in supervision.

Intelligence Tests and the Classroom Teacher. Arthur W. Kallom. Journal of Educational Research, 1922, May, 389–399. Practical suggestions for the use of intelligence tests by class-room teachers, illustrated by actual data and individual cases.

Some Pitfalls in the Administrative Use of Intelligence Tests. M. R. Trabue. Journal of Educational Research, 1922, June, 1–11. Cites errors made when tests are used by administrators who have no special training or experience in the field of measurement. Shows need for psychologist and administrator to work hand in hand.

How Much Mental Ability Does a Teacher Need? W. B. Bliss. Journal of Educational Psychology, 1922, June, 33–41. Data on relation between mental ratings and success in teaching.

Comparison of the Binet-Simon and Otis Tests. S. C. Garrison and J. S. Tippet. Journal of Educational Research, 1922, June, 42–48. A study of 158 pupils of the Peabody Demonstration School. Six tables present data.

Psychological Examination of Pre-school Age Children. David Mitchell. School and Society, 1922, May 20, 561–568. A study of 1113 pre-school age children of New York City shows the value of a preliminary examination and the need for special classification and modified curriculum. The examinations were conducted by the New York State Association of Consulting Psychologists.

Intelligence Tests and Collegiate Selection. Dagny Sunne. 1922, May 27, 593–595. Experiments with various tests at Newcomb College show that intelligence tests do differentiate students rather definitely, especially in the freshman year.

Mental Tests and College Teaching. Wm. R. Wilson. School and Society, 1922, June 10, 629–635. Discusses lack of co-relation between ability of college students, as shown by mental tests and their college grades. Suggests ways of

making each student work up to the limit of his powers—especially those of the superior group.

The Normal Curve and the Distribution of Intelligence Ratings. Garry C. Myers. School and Society, 1922. June 17, 676–678. Shows distribution in curve form of 5115 scores, Grade I to college inclusive, on Myers Mental Measure intelligence scale.

Sectional Differences as Shown by Academic Ratings and Army Tests. Martha McLear. School and Society, 1922. June 17, 676–678. A study of Northern and Southern negroes on basis of academic standing and army tests. Negligible difference in academic standing. Wide difference in mentality. Northern negroes higher.

Note on a Method for Studying Causes of Increase in Alpha Scores. Margaret V. Cobb and H. A. Tape. School and Society, 1922, June 24, 706–708. A study of the increase in Alpha scores of 81 pupils on three successive examinations comparing the increase from one class to the next, and the increase from year to year of each class group.

The Results of the Thorndike Intelligence Examination in the Senior Class of the Horace Mann School for Girls. Clara F. Chassell. School and Society, 1922, May 6, 511–512. Good evidence as to the fitness of Horace Mann graduates to do college work. The range of scores of the 54 students tested was from 41 to 94. Sixty-six per cent qualified absolutely, 28 per cent can probably do college work if specially industrious. Six per cent would be likely to prove unsuitable material for colleges and universities of high standards.

Eliminating First-grade Failures through the Control of Intellectual, Physical and Emotional Factors. Grace Arthur. School and Society, 1922, Apr. 29, 474–489. A study of 36 Grade I children and their progress during the first year of school. Case studies describe the actual work done in the ungraded room.

The General Philosophy of Grading and Promotion in Relation to Intelligence Testing. Henry W. Holmes. School and Society, 1922, Apr. 29, 457–461. Points out the need for segregation of gifted children and an enriched curriculum for them. Discusses the dangers of rapid advancement.

Shall We Classify Pupils by Intelligence Tests? Frederick S. Breed. School and Society, 1922, Apr. 15, 406–409. The difficulties involved in classification by intelligence tests. Discussion and practical suggestion.

The Discriminative Value of the Sub-tests of a Group Intelligence Test. Dora K. Mohlman. School and Society, 1922, Apr. 8, 399–400. Study of the results secured by administering the Indiana Group Scale of Intelligence to 77 university juniors and seniors. Correlation coefficients are reported for each part of the scale with the total score.

Intelligence as a Factor in the Election of High School Subjects. S. R. Powers. The School Review, 1922, June, 452–455. A comparison of the subjects elected by the high school students of Fort Smith, Arkansas and their scores on the Otis Intelligence Test. Students with high scores choose subjects making largest intellectual demands.

Intelligence Tests as a Basis for Homogeneous Grouping. May M. Harper. Elementary School Journal, 1922, June, 781–782. An experiment in the McKinley Junior High School, Xenia, Ohio, proves that classification on the basis of group intelligence test scores does place pupils fairly accurately.

How Different Mental Tests Agree in Rating Children. W. S. Guiler. Elementary School Journal, 1922, June, 734–744. A comparison of the Stanford-Binet, National Intelligence, Illinois Examination, and Pintner Non-Language Tests on the basis of IQ. Results of this study are invalidated by the fact that IQs computed on other tests are not comparable to the Binet IQ.

Educational Determinism, or Democracy and the IQ. Wm. C. Bagley. Educational Administration and Supervision, 1922, May, 257–272. An attack upon intelligence testing as inimical to the ideals and purposes of democratic education.

Intelligence and Behavior. A. A. Roback. Psychological Review, 1922, January, 54–62. A criticism of the behavioristic interpretation of intelligence. Reference to the definitions of intelligence in Intelligence and Its Measurement—Symposium. Journal of Educational Psychology, 1921, Vol. XII, p. 124.

A Comparison of Mental Abilities of Mixed- and Full-blood Indians on a Basis of Education. Thomas R. Garth. 1922, May, 221–236. Under conditions of uncontrolled social status but controlled school training the mixed-blood Indians surpass the full-bloods. Details and discussion of tests used.

Mental Tests and Mentality. T. H. Pear. Psyche, 1922, April, 304–314. Raises questions as to reliability of mental tests. Suggests paying more attention to "mental apparatus" and "mental attitude."

Intelligence Tests for Prospective Freshmen. Walter Dill Scott. Chicago School Journal, 1922, May, 321–324. A plea for vocational and educational guidance in college through an adequate Personnel Department.

The Revised and Extended Binet-Simon Tests, Applied to the Japanese Children. Y. Kubo. Pedagogical Seminary, 1922, June, 187–194. A revision of the Binet-Simon scale adapted to Japanese children. Performance tests and parts of the Otis and Army tests have been included. Range of ages—2 to 14 years.

Educational Tests

Instruments for Measuring Disciplinary Values of Studies. E. L. Thorndike. Journal of Educational Research, 1922, April, 269–279. A description of a test, made up in two series, *A* and *B*, for measuring the general improvement in generalization, relating, selection and organization as produced in a pupil by the study of grammar, languages, or mathematics.

Scales for Measuring Results of Physics Teaching. Harold L. Camp. Journal of Educational Research, 1922, May, 400–405. Description of the development of three scales designed to measure ability in (1) mechanics, (2) heat, (3) electricity and magnetism. Illustrative exercises and tentative norms are given.

Convenience and Uniformity in Reporting Norms for School Tests. J. Crosby Chapman. Journal of Educational Research, 1922, May, 406–420. Presents a scheme of reporting test scores by setting up nine equally separated levels of achievement for each grade. Criticises McCall's *T.* scale as a uniform procedure.

A Study of Reading and Spelling with Special Reference to Disability. Arthur I. Gates. Journal of Educational Research, 1922, June, 12, 24. Details of a study of the reading and spelling abilities of 135 children. Discussion of various defects associated with disability in these subjects. Brief mention of remedial treatment.

The Efficiency Quotient as a Measure of Achievement. T. L. Torgerson. Journal

of Educational Research, 1922, June 25–32. Advocates the use of an achievement quotient found by dividing the pupil's point score by the grade standard, and an efficiency quotient found by dividing his achievement quotient by his intelligence quotient. Tables for quotients embracing all standard tests have been prepared.

Measuring the Pupils in a Large City School. Joseph S. Taylor. School and Society, 1922, July 1, 25–28. Report of tests conducted by Dr. McCall and a group of graduate students at Public School 107, New York City. Grades tested were III*A* to VI*A*, including an ungraded open-air class and a class of mental defectives. Details of the various scores computed—educational age, pedagogical rank, promotion quotient, achievement quotient, etc., are given.

A Simplified Method of Determining a Pupil's Score on Gray's Oral Reading Test. W. S. Monroe. School and Society, 1922, May 13, 538–539. Using the same zero point for all grades and making the interval between paragraphs 4 instead of 5 provides a simple method of computing and interpreting scores on the Gray test. Norms expressed in terms of the new scale are given.

A Geography Test for the Sixth, Seventh and Eighth Grades. C. A. Gregory and Peter L. Spencer. School and Society, 1922, April 22, 452–456. A complete description of a comprehensive geography test, including discussion of criteria for choice of questions, and directions for giving and scoring the test. There are three duplicate forms each consisting of six parts.

Forecasting Failures in College Classes. Harvey B. Lemon. The School Review, 1922, May, 382–387. Description of a brief test used in the administration of undergraduate work in general physics at the University of Chicago. Elimination of probable failures and establishment of sympathetic acquaintance two of the beneficial results noted.

The Accomplishment Quotient—Finding and Using It. Katherine Murdoch. Teachers College Record, 1922, May, 229–239. A study of 415 children, Grades III to VIII in a large private school in Honolulu, Hawaii. Tests used were National Intelligence, Thorndike-McCall Reading, and Woody-McCall Mixed Fundamentals. Full discussion of process of obtaining accomplishment quotients and suggestions for the use of results.

The Cleveland Survey Arithmetic Test in Grade V-*B in Chicago.* Edw. E. Keener. Chicago School Journal, 1922, May, 336–344. Complete data secured from the use of the Cleveland test in 266 schools. Discussion of standardized tests as aids in teaching.

Miscellaneous

An Aid to the Analysis of Vocational Interests. J. B. Miner. Journal of Educational Research, 1922, April, 311–323. Description of an individual analysis blank which aims to train pupils to analyze their work interests. Two new features are a classification of occupations according to activities emphasized and a presentation of contrasts in working conditions.

An Accurate Index of Nationality. Riverda H. Jordan. Journal of Educational Research, 1922, May, 421–425. Argues that in studying the nationality of school children data should be based upon the birthplaces of grandparents rather than fathers.

The Psychology of Learning Applied to Typewriting. E. W. Barnhart. The American Shorthand Teacher, 1921, November, 1922, February. Four articles on the application of the well-established laws of learning to typewriting.

A Data Sheet for the Pearson Correlation Coefficient. L. L. Thurstone. Journal of Educational Research, 1922, June, 49–56. A labor-saving device for the calculation of the Pearson *r*. Illustration of data sheet and complete instructions for its use.

A Critique of Mental Measurements. Thomas J. McCormack. School and Society, 1922, June 24, 686–692. Expresses complete lack of belief in the whole science of mental measurement.

A New IQ Slide Rule. Lloyd N. Yepsew. School and Society, 1922, May 27, 596. Description of a circular rule device designed especially for use in computing intelligence quotients.

Classification in Athletics for the Purpose of Individual Self-rating. Jesse Feiring Williams and Myrtle Hummer. Teachers College Record, 1922, May, 240–254. Class divisions based on the actual accomplishment records of 1612 boys and 1773 girls of Trenton, N. J., permit self-rating in athletics and physical ability. Eleven interesting charts present details.

Variation in Grading High School Pupils. J. E. Armstrong. Chicago School Journal, 1922, May, 346–348. A criticism of our present grading system.

Vocational Interests of High School Seniors. Aubrey A. Douglass. School and Society, 1922, July 15. Analysis of the answers of 1658 girls and 1186 boys—all high-school seniors in the State of Washington—to a vocational interests questionnaire.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

1. *Mental Growth Curve of Normal and Superior Children Studied by Means of Consecutive Intelligence Examinations.*[1]

Dr. Baldwin was inspired with a fine idea when he conceived the plan of measuring at successive intervals the intelligence of a fairly large number of children. The Stanford-Binet examination was given to 56 subjects twice, to 51 three times, to 44 four times and to 36 subjects five times. In all 143 individual record cards were used. The excellent plan of computing the mental age for each exact chronological age was also followed. According to the authors (1) "The mental growth curve reveals a significant change in the trend with the approach of adolescence, which appears earlier in the case of superior children. There is also an adolescent superiority of girls which is in accordance with other facts indicative of the earlier maturity of girls;" (2) "the mental growth curves are strikingly similar to the physical growth curves in height;" (3) "the IQ curves are approximately horizontal, confirming within limitations the constancy of the IQ. There are fluctuations associated with physical development;" and (4) "the mean IQ of each of the four groups of children increased with each successive examination, which is probably an effect of greater habituation or practice."

So far, so good! If one accepts Dr. Baldwin's method of plotting the curves one can accept his conclusions. But hasn't he made an error in logic in this part of his task? To show mental growth he plots mental ages against chronological ages, and apparently is astonished when both the superior and the inferior group curves are roughly straight lines. What else could be expected? His method of plotting (Chart II) simply proves what other studies have also proved, namely, that the IQ from 5 to 14 is fairly constant. That this criticism is not mere quibbling is shown by comparing Chart II (a mental growth curve) with Chart IV (an IQ curve). They are the same thing

[1] Baldwin, Bird T., and Stecher, Lorle I.: University of Iowa Studies, Vol. II, No. 1, January, 1922, pp. 61.

plotted in different fashion, although diametrically opposite conclusions are drawn from them. (See conclusions (1) and (3) above.)

Isn't Dr. Baldwin thinking in a circle when he uses mental age as a measure of mental growth? Isn't he assuming that the mental growth from 5 to 6 is the same as the mental growth from 12 to 13? It may be that it is so, but no one has yet proved it. Until an absolute unit for measuring intelligence is devised it will be impossible to say that the curve of growth of mental age is similar to the curve of growth for height, or to any other curve for that matter. Theoretically, the curve should be logarithmic. And the records from a large number of group tests give indications of a logarithmic character. But until the absolute unit has been discovered such researches as this are quite beside the mark. What Dr. Baldwin has done is a good piece of work on the question of the constancy of the IQ. He shows conclusively that it is fairly constant though subject to fluctuations so far as the measurements made are reliable. He has also pointed out that the results of repeated examinations exhibit a definite practice effect. But he has *not* plotted a mental growth curve although it pains the reviewer to have to point this out. But if it will soothe Dr. Baldwin's feelings the writer confesses that he (as well as many others) has previously stumbled into this very error.

PETER SANDIFORD.

---

2. *A Study of Superior Children.*—The gradually increasing number of studies dealing with superior children shows that psychologists and educators are beginning to realize the importance of knowing more about this type of child. The author of the monograph under consideration[1] presents in sociological and psychological study of a small group of children of superior intelligence. The children were selected from among those reported by principals and teachers as superior. The younger children had IQs of 135 or above, and the older had IQs of 120 or above. One or two cases with lower IQs were studied, the lowest being 117. These children were given a great number of tests of all kinds, opposites, symbol-digit, directions, proverbs, and the like. The superior children excelled normal children on all these tests.

The sociological part of the study includes very elaborate case histories of each child. These are very interesting in as much as we

[1] Root, W. T.: A Socio-Psychological Study of 53 Supernormal Children. *Psychological Monographs*, Vol. 29, No. 4. Whole No. 133. Princeton, 1921.

have little of this sort of data as compared with the amount we possess for subnormals. In general the superior child is characterized as having a good home and superior parents. The author then attempts to explain his results and proceeds to a long discussion of the central common factor theory. According to our author the common factor could just as well be environmental as innate. "The common factor or factors may be a varying admixture of innate ability, formal training, incidental education and social conditions." Indeed, all through the thesis the author emphasises the importance of environmental factors much more than the average psychologist would, and one is inclined to question the soundness of some of his opinions in this respect. It is, however, well to have this side of the picture presented, even although we cannot agree. There seems, furthermore, a feeling of dissatisfaction on the part of the author with the present intelligence tests or with tests in general. He feels that no tests measure the real basis of superior intelligence, for this real basis consists of ability to suspend judgment, freedom from suggestibility, critical attitude, etc. Again there is room for much argument over these phrases. The results presented in the monograph are important and interesting. The conclusions and the opinions of the author are open to much debate.

R. P.

---

3. *Group Intelligence Tests in England.*[1]—This book gives American psychologists the first account of group testing in England and we have in its author, Mr. Ballard, a most delightful cicerone for our tour, one who knows well how to mix humor with his learning. We must remember, however, that he is not acting as guide to Americans, but is explaining the field to English teachers. So much the more interesting, therefore, is it to the American psychologist to hear the explanation of our own group tests. Here is their origin. "Individnal testing was born in France; group testing was born in America. And its mother was necessity—the stern necessity of war." And so he tells of the testing in the army. "The whole undertaking was a colossal business; and the official report which has recently been issued is correspondingly colossal. It weighs about four pounds." Mr.

[1] Ballard, P. B.: Group Tests of Intelligence. Hodder and Stoughton, London, 1922, pp. X, 252.

Ballard admires our fool-proof tests, and he says very aptly that "a fool-proof test is one that prevents the examiner from making a fool of himself." And so he ambles along with a delightfully humorous, but thoroughly sympathetic, account of the work in this country. It is an account that no psychologist should fail to read.

For American readers the description of group testing in England is of particular interest. Although little has so far been done, the use made of intelligence tests by the Bradford and Northumberland Education Committees for the selection of scholarship children is important and significant, and one feels that in the course of time the extent of the work must inevitably increase. Our author also gives samples of several group tests used in England. Many of these are clearly descendents of the American group test, but there are innovations and new ideas that are very suggestive. The English workers are not quite as partial to short time limits as most American workers. New types of material for group tests are Thomson's Hindustani Test and Ballard's Cipher Test, Orientation Test and Cryptogram Test.

The book also contains a very simple chapter on the nature of intelligence in which the various definitions of intelligence are discussed. It is admirably written and forms a decided addition to the literature of the subject. It ends with the author's own definition as follows, "Intelligence is the relative general efficiency of minds measured under similar conditions of knowledge, interest and habituation."

Other chapters in the book deal in a simple manner with correlation and statistical procedure, and with the use of group tests in schools. The last chapter in the book has no connection with the book proper. It is on "Spelling Demons." The reviewer wonders why the author included this unrelated topic. Perhaps as a mental test for reviewers to see if they could discover the relationship of this small part to the whole book. If so, the present reviewer has failed.

R. P.

---

4. *Tests of College Students.*—The college student has been the most popular "research animal" in the psychological laboratory, and the present monograph[1] is another addition to the long list

[1] Carothers, F. E.: Psychological Examinations of College Students. *Archives of Psychology*, No. 46, December, 1921.

of studies reporting psychological tests of college students. It contains a very good summary of the previous work done in this connection. The contribution of the author herself is a detailed study of 19 different tests given to various groups of college freshmen. The results show the possibility of finding "several groups of tests which correlate closely among themselves, but loosely with the other tests." There is no evidence in support of Spearman's theory of a general common factor in so far as this depends upon the inter-correlations.

The author's attempt to explain the correlation, or the lack of it, of the separate tests with various college subjects is very naive. It is hard from a study of the actual coefficients to believe, as the author tells us, that "the five academic groups show positive correlation with tests which we would expect to correlate with them." Mathematics shows the highest correlations with Cancellation, Checking, and Knox Cube; Science with Opposites, Verb-objects, Mixed Relations, Knox Cube, and Logical Recollection. An attempt to explain why is made by the author on the basis of the mental processes involved. Fortunately she refrains from telling us why Philosophy should correlate highest with Cancellation, Word Naming, Knox Cube, and Digit Span.

The practical value of psychological tests for students' guidance is stressed. A psychographic chart for each student is recommended, which will show respective strengths and weaknesses.

The work is carefully done and is a valuable addition to our knowledge of the psychology of the college freshman. R. P.

---

5. *Juvenile Delinquency.*—The book under consideration[1] presents in interesting form a report of the work being carried on at The Ohio Bureau of Juvenile Research. As such it is of value to educators, who have to deal with disciplinary problems, and with the questions arising from the presence of mental deviates in schools.

A special chapter is devoted to the psychopathic child, who may be of any degree of intelligence, but who shows abnormalities of mental functioning apart from capacity for learning. A state institution for psychopathic children is recommended.

The author believes that "juvenile delinquency can be largely

[1] Goddard, H. H.: "Juvenile Delinquency." Dodd, Mead and Company, New York, 1921, pp. 120.

eradicated" if the states are willing to pay the price of studying and caring for unfortunate children. Research should be one of the state's chief concerns in the present condition of knowledge, for as yet many fundamental questions relating to the origin, development and modifiability of unfortunate deviates cannot be answered.

LETA S. HOLLINGWORTH.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Vol. XIII October, 1922 No. 7

QUOTIENTS I, E, AND A

RENA STEBBINS

AND

L. A. PECHSTEIN

City Normal School, Rochester, N. Y., Department of Psychology and Education, University of Rochester

When an examiner has finished giving a battery of educational tests in a modern school, he has two groups of data to offer the principal—first, a record of grade norms in comparison with standard norms, showing what grades have central tendencies above, equal to, or below standard; second, individual class record sheets for each test, showing what score each child made on each test. And as the examiner gives this material to the teacher, to the supervisor, or to the principal, he is struck with the unfairness of it all—its unfairness to the teacher, because the records show only how her grade compares in measurable educational abilities with other grades in the country, without taking into account the pupil-material with which she worked; its unfairness to the pupil, because it gives high rank and praise to those above the norm and a sense of failure and blame to those below the norm, without taking into account the native capacity or intelligence of these children.

Those interested in educational testing have long realized that the valuable and fair testing program will be one in which the results of educational tests are compared with the results of mental tests; in which we measure both teacher and pupil not by what the child has learned, but by what he has learned in relation to what he is capable of learning. The need for a combination of educational and mental tests has long been felt. The method has been the chief difficulty. Several different lines of approach have been taken in various parts of the country. In Rochester, we have chosen the method which follows because of its simplicity, its fairness, and its possibilities for interpretation.

The School Situation

Four hundred children in four grades in the Practice Department of the City Normal School, Rochester, N. Y., were chosen for our field of testing. Because of other testing projects in the first three grades, we limited our work to Grades IV through VII. Each year of school work is divided in two, and promotions are half-yearly. The first half-year in a grade is designated by B, the second half-year by A. Not only are there B and A grades for each year, but in every grade there are two parallel classes, one composed of superior children, one of inferior children, classified on the basis of teachers' judgments. Thus there are in the fourth grade:

IV B_1—First half of Grade IV, brighter pupils
IV B_2—First half of Grade IV, slower pupils
IV A_1—Second half of Grade IV, brighter pupils
IV A_2—Second half of Grade IV, slower pupils

In the four grades studied there were in all, then, 16 classes, taught by 14 well-trained teachers.

Selection and Administration of Tests

In selecting tests for our educational survey, no attempt was made to choose new tests for the sake of evaluating them. Only well-known, widely used tests were selected, tests whose value had already been clearly demonstrated elsewhere. The tests chosen were:

Mental
- National Intelligence Test A and B, Form I.

Educational
- Reading Thorndike-McCall
- Arithmetic Fundamentals: Woody Addition Scale A; Woody Subtraction Scale A; Woody Multiplication Scale A; Woody Division Scale A
- Problems Buckingham Problem Scale
- Spelling: Ayres' (in lists); Monroe Timed—Sentence Spelling Test.

The group mental tests were given by Miss Leila Martin, Director of the Child Study Bureau of Rochester, under excellent and uniform conditions. The educational tests were all given by one trained examiner under uniform testing conditions, and all corrections, tabulations and statistical work were done either by this examiner or

under her personal supervision. Because of this great care in the administration of the tests, we feel that mechanical error has been reduced to a minimum and that conditions warrant a close comparison between grades.

The Scores

When the tests were given, we had scores in two group intelligence tests and in seven educational tests for each child. The typical record for each child follows:

A. J.—Age 10 Years 3 Months	Score
National Intelligence Scale A	94.0
National Intelligence Scale B	96.0
Thorndike-McCall Reading Test	46.0
Woody Addition Scale A	28.0
Woody Subtraction Scale A	23.0
Woody Multiplication Scale A	27.0
Woody Division Scale A	24.0
Buckingham Problem Scale	6.46
Ayres' Spelling Scale	77.0
Monroe Timed—Sentence Spelling Test	42.0

One glance at this record makes the difficulty apparent. When these scores are given to the teacher, they are valuable only insofar as she may compare them with scores of other children in her grade. This comparison she had already made, however, long before we gave the tests. A more concrete comparison she cannot make. If these scores should ever be given to a parent, he would get almost no information from them, except the judgment that he, who used to be "good in reading" has a child who "stands 46." Even the examiner himself cannot interpret or compare results based upon such varying criteria without further work. Clearly before we can study the relation of this child's educational achievement to his mental capacity, we must reduce these unrelated scores to some comparable unit. The method of composite scores, and the method of composite ranks seem to yield results less capable of interpretation to the average teacher and parent than the method of the educational age.

Mental and Educational Ages

In the field of intelligence tests, all efforts to substitute scores, percentiles, intelligence indices, etc., for Binet's and Terman's respective contributions of the mental age (MA) and the intelligence quo-

tient (IQ) have failed. The layman understands MA and he wishes the results of the intelligence test stated in terms which he can comprehend. In the field of educational measures, if we could give both the teacher and parent in the case of the child mentioned above, instead of a group of scores, a statement like this: "The child reads as well as the average child of 11 years 6 months; in addition, he does as well as the average child of 11 years 10 months; in subtraction, 11 years 6 months; in multiplication, 12 years; in division, 11 years 10 months; in problems, 12 years; in list spelling, 12 years 4 months," we would make test results immeasurably more valuable to teacher, pupil, and parent. This is what we have attempted to do with the scores obtained in our study.

The transforming of national intelligence scores into mental ages was simple. Using the age norms given by the National Research Council, we interpolated for months, forming an age-score table as follows:

Provisional Age Standards
National Intelligence Test Form A-1

Years	Months	Score	Years	Months	Score	Years	Months	Score	Years	Months	Score
8	0	65.0	9	0	78.0	10	0	91	11	0	103.0
8	1	66.1	9	1	79.1	10	1	92	11	1	103.8
8	2	67.2	9	2	80.2	10	2	93	11	2	104.6
8	3	68.3	9	3	81.3	10	3	94	11	3	105.5
etc.			etc.			etc.			etc.		

From these tables we read the mental ages which corresponded to the scores on Scales A and B, and averaged them to find the child's mental age. While no one experienced in the inaccuracies of group testing at its best would care to state that this is the child's true mental age, at least it is the mental age which corresponds to his scores obtained on group tests given under the best possible conditions.

Our next task was to transform educational scores into what we have called an educational age. If we say that a child has a mental age of 12 years when he does as well on a mental test as the average child of that age, then we may say that a child has an "educational age" of 12 years when he does as well on educational tests as the average child of that age. It is clear that in order to transpose scores into educational ages the same method of standardization is necessary that is used in finding age norms for mental tests. Large unselected

groups of children at each age, taken regardless of the grade in which they are located, should be given the test, and age standards found from the median or mean performance of these children. This has been done in the case of the Thorndike-McCall Reading Scales. In a clear discussion of scale construction Dr. McCall[1] points out how he procured the educational age norms for his 10 reading scales. The examiner, by using the table for reading age standards sent with the manual of directions, can find directly and accurately the "reading age" for each child. This has been done with all our reading scores.

Method of Deriving Educational Ages

For the other tests in our study, only grade norms were furnished by the test originators. Our problem was to find some method of using these grade norms given with the tests to find educational ages. We realize that the method we have used has its defects; that it is at best only a makeshift; that it is as accurate as it could now be made and, until test constructors furnish us with age norms, it is the best method for interpreting scores.

Let us take for an example the Buckingham problem test. Dr. Buckingham has found the following grade medians for the end of the term, based on large numbers of pupils tested:

End of	Score
III*B*	3.66
III*A*	3.84
IV*B*	4.36
IV*A*	4.74
etc.	

It is generally accepted by educators, and the 1918 school census reports support the facts, that the standard median ages for the above grades are:

End of	Years	Months
III*B*	9	
III*A*	9	6
IV*B*	10	
IV*A*	10	6

We have said, since the median performance in the problem scale for III*B* grade is 3.66, and since the median age for III*B* grade is 9

[1] McCall, William A.: Uniform Method of Scale Construction, *Teachers College Record*, Jan., 1921.

years, we shall let 9 years be the "problem age" for all children making a score of 3.66 on the scale. Hence we derive the following table:

Age Standards for Buckingham Scale

Score	Age		Score	Age	
	Years	Months		Years	Months
3.66	9	0	6.46	12	0
3.84	9	6	6.61	12	6
4.36	10	0	7.89	13	0
4.74	10	6	8.15	13	6
5.86	11	0	8.27	14	0
6.13	11	6	8.56	14	6

By interpolating for months, we found the complete table, and from this table we read our scores in terms of problem ages. This same method was used in finding age tables for each scale used. With the spelling scales, we were limited by the fact that the scales do not run continuously through the grades. In the Ayres' list it was necessary to make a separate table for each column, using Ayres' standard per cents for each grade. In Monroe's timed-sentence test, we made a separate table for the test for Grades III and IV, for Grades V and VI, and for Grades VII and VIII.

When we had secured an educational age for each child in each subject tested, we wished to find the arithmetic mean of these ages to procure an average educational age. There were, however, four tests in arithmetic processes, and two in spelling, as against one test in reading and one in problems. We wished our mean to give equal value to skills in each subject. It would be necessary, therefore, to weight the reading age by 4 as a multiplier, the Woody ages by 1, the problem age by 4, and the spelling ages by 2 each in order to equalize their value. This was done in every case except in the two spelling ages. Although we considered spelling ability as important as the other abilities, still we felt that our spelling ages were not reliable enough to warrant equal weighting with the others, and we decided to weight each spelling age by 1 only, thus giving only half value to spelling. Mean educational age, then, is an average of the educational ages in each subject in the following proportion:

4 ×	Reading age	=
	Addition age	=
	Subtraction age	=
	Multiplication age	=
	Division age	=
4 ×	Problem age	=
	List spelling age	=
	Sentence spelling age	=
	Sum of EA's	=
	Divided by 14	= = Mean EA

Educational Ages in C. N. S.

Table I illustrates the record of the chronological, mental, and educational ages which was made for each of the 16 grades. A similar tabulation was put in the hands of the teacher of each grade for reference during promotion week, as summing up for her in an intelligible manner, the results of the educational tests given to her pupils. Compared with the usual method of representing test results, we feel that this summary is far more valuable for diagnosis of individual differences than the usual "list of scores in order" method.

Table I.—A Table Showing the Chronological, Mental, and Educational Ages for Pupils in a Typical VI*B* Grade

Pupils, *girls*	CA	MA	Educational ages								Mean EA
			Reading	Addition	Subtraction	Multiplication	Division	Problem	Sentence spelling	List spelling	
Z. C.	11-3	10-3	11-3	11-4	12-7	14-2	13-5	12	12-0	12-6	12-1
C. C.	12-8	10-4	12-8	11-8	14-0	13-2	12-0	11-0	12-6	12-10	12-2
C. D.	12-10	11-5	11-3	11-10	12-4	13-5	11-10	11-11	12-9	13-0	12-0
D. D.	11-6	11-2	11-3	12-8	14-2	14-2	14-4	12-10	12-9	12-8	11-8
M. E.	10-11	11-9	11-9	11-10	12-10	12-5	14-4	12-6	13-0	13-0	12-6
O. F.	12-3	9-5	13-2	12-0	12-0	13-2	11-8	11-11	10-0	11-0	12-3
S. I.	11-4	10-9	12-8	12-8	13-8	11-10	11-8	11-11	11-6	11-6	12-3
R. K.	10-11	12-5	10-10	14-5	14-0	13-11	15-0	12-6	13-4	13-4	12-8
W. M.	11-4	9-5	11-9	14-9	12-4	13-2	14-4	10-1	11-5	11-10	11-10
G. S.	10-10	11-0	12-8	11-8	12-0	13-11	11-10	11-11	12-2	12-4	12-4
B. W.	12-4	13-6	11-3	13-0	13-8	13-2	11-8	12-6	12-9	12-2	12-3
Median..	11-11	11-1	.....	.....	.....	.....	.....	.....	.....	.....	12-1

With our material now in terms of comparable units, we were able to combine the mental ages and educational ages in one measure—the

Accomplishment Quotient. The Accomplishment Quotient is the ratio of a child's educational age to his mental age; it is the summary of what a child accomplishes educationally compared with what he is capable of accomplishing. In the case of a 10-year-old child who is 12 years mentally, but who does as well in his work as the average 11-year child, his IQ is $^{12}/_{10}$ or 120; his EQ is $^{11}/_{10}$ or 110, but his AQ is $^{11}/_{12}$ or 92. His IQ of 120 shows him to have superior intelligence for his age, but it does not show what he is doing with that intelligence. His EQ of 110 shows that he is doing better work than average for his age, but it does not show that he is doing as well as he is capable of doing. His AQ tells the whole truth—he is accomplishing only 92 per cent as much as he would normally be expected to accomplish, considering his intelligence. We believe that this Accomplishment Quotient is the fairest and most valuable measure now known of the efficiency both of the pupil and the teacher.

The Accomplishment Quotient and the Teacher

From the point of view of the teacher, the Accomplishment Quotient is the only quotient which takes into account the material with which she is working.

1. For the teacher of slow pupils, the AQ interprets fairly the results of her work in relation to the capacities of the children. Her *low* educational scores may appear high when we compare attainment with native capacity.

2. For the teacher of bright pupils, the AQ interprets the educational results fairly. Her *high* scores may not be as high as they should be, when compared with the native ability of the children whom she teaches.

3. For every teacher and every supervisor who is supplied with the AQ's at the beginning of the term, this measure is a protection against injustice. If the class has been improperly taught for 2 or 3 years, the AQ reveals to the teacher and to the supervisors the handicap under which the new teacher is working. If the class has been unusually well taught for 2 or 3 years, the AQ reveals the fact and prevents the new teacher from resting too complacently upon the past laurels of the class.

An illustration from one grade (VIA_2), will demonstrate the unfairness of the "score" method of showing results, and the fairness of our proposed Quotient Method. If we represent the grade medians in

FIG 1.—Results of eight educational tests, Grade VIA_2 when judged by:
Grade medians as per cents of norms (——).
Accomplishment quotients (. . . .).

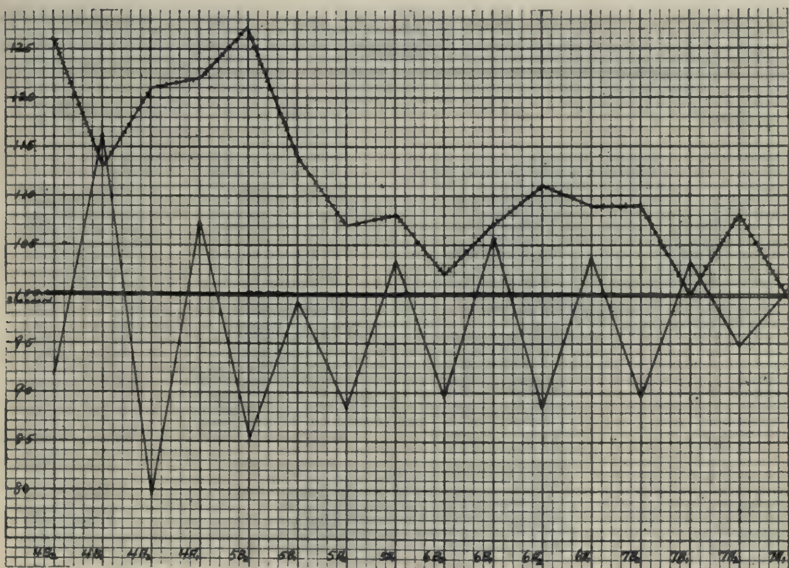


FIG. 2.—Comparison of educational test results in all grades when judged by:
Average grade medians as percepts of standards (——).
Average accomplishment quotients (. . . .).

each test as per cents of standard scores (See Figure 1), the teacher of this grade faces discouraging results. In every test her grade is below the norm. In only three tests is her grade above 95 per cent of the norm. The teacher and the principal realize that the VA_2 grade is a slow grade, but the record does not acknowledge that fact. If however, we represent these results not as per cents of norms, but in terms of median Accomplishment Quotients in each test (*i.e.*, the median of the educational ages in each test divided by the median metal age), notice how this grade shifts. In every test this grade is doing as well as or better than might be expected from its mentality. This teacher should not be discouraged, but encouraged by the excellent progress her children are making in skill subjects in the light of their mentality.

The facts of a grade are consistent for the entire school. Figure 2 shows the average results for each grade. The lower line represents averages of the grade medians divided by the grade norms. In every case, the duller section of the grade presents a discouraging showing. All but one of the brighter sections are above standard in these subjects. But when we take into consideration the mentality of these grades, and represent the results in terms of average

Table II.—Table Showing Median IQ, EQ, and AQ for all Grades in Rank Order

Rank	IQ	EQ	AQ
1	$VIIB_1$—103	IVB_1—107	VB_2—127
2	$VIIA_1$—101	VIA_1—105	IVB_2—126
3	VIA_1— 98	$VIIB_1$—103	IVA_1—122
4	VB_1— 97	VB_1—102	IVA_2—121
5	IVB_1— 96	IVA_1—102	VB_1—114
6	VIB_1— 94	$VIIA_1$—101	IVB_1—113
7	VA_1— 94	VA_1—101	VIA_2—111
8	VIB_2— 93	VIB_1—100	$VIIB_2$—109
9	VA_2— 90	VA_2— 95	VIA_1—109
10	IVA_1— 85	VIB_2— 94	$VIIA_2$—108
11	$VIIA_2$— 84	VB_2— 92	VA_1—108
12	$VIIB_2$— 84	$VIIA_2$— 90	VIB_1—107
13	VIA_2— 81	$VIIB_2$— 89	VA_2—107
14	IVA_2— 74	VIA_2— 89	VIB_2—102
15	IVB_2— 72	IVA_2— 89	$VIIA_1$—100
16	VB_2— 69	IVB_2— 99	$VIIB_1$—100

Accomplishment Quotients, as shown by the upper line in Figure 2, we find that the teachers who seemingly are doing the best work are those in charge of the slower groups. Two dull sections are achieving over 125 per cent of what might be expected from them. The IVA_2 and the VB_2 grades, the two grades whose medians were lowest in the school, are accomplishing 121 per cent and 127 per cent as much as might be expected from their mental ability and this is the true measure of a grades' achievement—not how much can it score, but how much can it achieve in relation to its ability?

Interesting facts revealed by Figure 2 are that the lower grades seem to be more efficiently taught in these skill subjects than the upper grades, and that all grades in the school are accomplishing as much or more than might normally be expected from their mental abilities.

Table II represents a summary of the median IQ's, EQ's, and AQ's of the 16 grades in order of their rank. The brightest grades tend to have the highest educational age but they tend, after all, to be the laggards, for it is the duller grades that have relatively superior performance. While our Accomplishment Quotients are not an entirely just measure of the teacher's efficiency in these eight phases this year, since they do not take into account with what quotients the grade came to her, they are valuable in revealing the results of 1 year under her teaching, and in another year, we feel a similar record will show very fairly the teaching efficiency of every teacher in these skill subjects.

The Pupil and the Accomplishment Quotient

Educational test results taken alone are no fairer to the child than to the teacher. The bright child receives the high score and the praise; the duller child takes the low score and defeat, with no regard given to the comparative mentalities. The Accomplishment Quotient is a just measure of the pupil's efficiency in school work.

1. For the bright children, it shows which child is living up to his possibilities and which child fails to make his attainment equal to his capacity to attain.

2. For the dull children, it shows which child is needing to be urged and helped still more, which child needs restraining, perhaps, and which are most deserving of praise.

3. Of all children, it asks that the pupil be urged to progress at a rate which is proportional to the mental capacity with which nature endowed him. This is the only fair standard for any child.

Of the 359 pupils tested in the 16 grades, the Accomplishment Quotients are distributed as follows:

Accomplishment Quotient	Number of Pupils
140–150	5
130–139	10
120–129	25
110–119	63
100–109	96
90–99	93
80–89	53
70–79	13
60–69	1
Total............	359

If we may state that, roughly classifying, an accomplishment quotient of from 90 to 110 is satisfactory, there are, then, about 189 pupils doing satisfactory work, about 103 doing better than might be expected from their native endowment, and about 67 who are apparently not fulfilling their educational possibilities. That is, 52 per cent are doing satisfactory work, 29 per cent are over-attaining, and 19 per cent are falling short of the demands the school may properly lay upon them.

The value of the Accomplishment Quotient in relation to the child is most clearly revealed when we study individual cases. The records of six Grade VII children are reproduced in Table III.

D. M. is an 82 IQ boy, 13 years 8 months of age, doing the work in these subjects of the average child of 12. Instead of censuring the boy for being backward for his age, we see he deserves praise for accomplishing 108 per cent as much as might be expected of him. His profile reveals the fact that his educational work is more than satisfactory in all but reading and spelling, where his educational age lags behind his mental age. Here he needs special attention.

F. Z. is a boy 12 years 8 months of age, with an average educational age of 13 years 6 months. He is doing entirely satisfactory work from the teacher's point of view except for inaccuracy in arithmetic. But his mental age is 16 years 8 months; his accomplishment quotient is only 80. F. Z. is a laggard. Instead of receiving credit for doing well, he needs to be stimulated to do far better work.

E. D. is a normal girl, with an IQ of 101, and an EQ of 99. She is doing satisfactory work for her ability, on the whole.

G. M. is a girl who has been kept after school all her life, and she

TABLE III.—A TABLE TO SHOW THE CHRONOLOGICAL, MENTAL, AND EDUCATIONAL AGES OF SIX VIIA PUPILS SELECTED FOR SHOWING THE INTERPRETATIVE VALUE FOR THE CHILD OF THE SEVERAL MEASURES PRESENTED

Pupil	CA	MA	Educational ages									Quotients		
			Reading	Addition	Subtraction	Multiplication	Division	Problem	Sentence spelling	List spelling	Mean EA	IQ	EQ	AQ
D. M.	13-8	11-2	10-10	12-4	13-3	13-2	13-11	12-10	10-10	10-3	12-0	82	88	108
F. Z.	12-8	16-8	13-11	11-8	12-7	11-5	11-10	14-11	13-9	12-6	13-6	132	107	80
E. D.	13-9	13-11	12-8	15-1	12-10	13-5	13-5	13-9	14-10	14-2	13-6	101	99	97
G. M.	16-9	9-6	11-3	11-0	11-9	11-7	12-0	12-11	10-10	9-9	11-8	57	70	123
S. R.	13-4	15-7	14-4	14-5	14-0	14-6	14-8	13-9	14-8	13-5	14-2	117	106	91
S. F.	10-4	16-0	15-6	15-5	14-4	14-6	14-4	15-6	15-2	16-0	15-3	155	148	95

has faced daily failure and discouragement. Will a teacher ever again scold this faithful child after seeing her accomplishment record? With a mental age of 9 years 6 months, this child is doing the work of an average 11 years 8 months pupil. What does her age of 16 years 9 months matter, when we know that she is accomplishing 123 per cent as much as we might reasonably expect.

S. R. is another case of a bright child, doing supposedly satisfactory work, but really working far beneath her maximum capacity. This girl has never been stirred to do her best; she has been one of the best in the class, and has been praised for doing 91 per cent of what she is capable.

S. F. is a brilliant boy whose unusual intelligence was clearly recognized 2 years ago. In that short time, his educational age has been brought up to 15 years, 3 months, although he is only 10 years 4 months old. He is still accomplishing only 95 per cent of what might be expected from his 16-year intelligence, but it is a fair question whether it is socially wise to give him the educational opportunities sufficient to allow him to make more rapid progress. His present AQ of 95 is a

tribute to what it is possible to do, once the situation is recognized by the teacher.

EA and MA in Classification

The study of individual cases revealed the great need for reclassification in many instances. There was great overlapping of mental ages in the brighter and duller sections of a grade, consequently a conservative reclassification was organized on the following principles:

1. A child whose MA is above the median MA of the next grade should be considered for double promotion.
2. Of these, only those children whose EA's are above the median EA of their present grade should be allowed a double promotion, since without the necessary tools for acquiring the work of the omitted half-year, the child could hardly hope to succeed.
3. A child whose MA is below the median MA of the grade below should be considered for non-promotion.
4. Of these, only those children whose EA's are below the median EA of their present grade should be non-promoted.
5. All other children should be regularly promoted, and assigned to the brighter or duller group on the basis of test results.

It is the opinion of the writers, however, that group test results must always be subjected to the judgment of the teacher for corroboration, and where disagreement is sufficiently marked to cause a possibility of injustice to the child, individual case study must follow. In no other way can we avoid letting cold figures work disastrously in the career of a living child.

In conclusion, may we give a statement of our faith? We believe that the Accomplishment Quotient is the fairest and most valuable measure of both the efficiency of the teacher and the pupil, that by reliance on it for guidance, the teacher will come to exact from him that hath even more than he has been giving and take from him that hath not even less than he has been able to give. And the educational plaudits of "well done" are seen to be merited more by the retardate possessing the one than by the accelerate endowed with the ten.

THE TEACHING OF EDUCATIONAL PSYCHOLOGY IN THE UNITED STATES

H. H. REMMERS

Colorado College

AND

F. B. KNIGHT

University of Iowa

No one any longer thinks, seriously or otherwise, of describing psychology, and particularly educational psychology, as "putting what everybody knows in language which nobody can understand."[1] The viewpoint of Pyle as expressed in his recent book is probably nearer the concensus of opinion among those qualified to judge in the matter. "Educational psychology," he says, "is an experimental science,"[2] and it will be our purpose so to treat it in what follows.

To determine the extent to which experiment is used in elementary courses of educational psychology has been the object of the study to be described in this chapter. What is the present practice? To what extent do the beginning courses in educational psychology—which represent the sum total of the training in psychology that all but a negligible part of young teachers in training receive—attempt to give the prospective young teachers a mastery of the subject through an experimental approach? Such a mastery, that is, that will give her the *savoir faire* which comes only from the ability to apply theory directly and accurately to classroom problems?

It might be contended, if it did not assume the reader's consent without argument, that the human mind induces a theory from much practice rather than that it deduces current practice from theory. An order of merit of learning techniques might be roughly:

Best—doing the thing oneself.
Next—seeing it done.
Third—reading about what happened when it was done.
Last—reading about the theory that underlies the operation.

Thus many a good teacher takes pains to make the assignments clear, exact, and definite without knowing that "ease of identification of

[1] Welton, J.: "The Psychology of Education," (1912), p. 1.
[2] Pyle, W. H.: "The Psychology of Learning," (1921), preface.

bonds" is helpful in learning; conversely, many a conscientious student of educational psychology knows in a verbal fashion that "ease of identification of bonds" is an aid to learning, but never sees that it means something in the practical situation of assigning lessons.

Further, what are considered, for the elementary student, to be the most profitable experiments? Answers to these questions were obtained through correspondence with the leading universities, colleges, normal, and teachers' training schools throughout the United States. The letter sent is here inserted:

DEAR SIR:

In a study that is being conducted in the Graduate College of the State University of Iowa, we are attempting to gain a true picture of the place of experiment in *elementary* or *first courses* of educational psychology.

The following meanings are attached to the terms as used: (1) An *experiment* is that which is done by the students themselves singly or in groups; (2) a *demonstration* gives the students an opportunity to observe an experiment as carried out by the instructor and assistants; (3) a *recountal* of experimental data places emphasis in lectures on experiments carried out by research workers.

We shall greatly appreciate your cooperation in this study, and will reciprocate by sending you a report of our findings if you wish it. Please place a check-mark (√) in the appropriate column after the items in the accompanying list thus indicating what you are offering in your courses. Also please answer the three questions at the end of the list.

Very truly yours,

——— ———.

Note: These experiments have purposely been left disorganized, since no one organization would correspond to the sequence of topics in all courses.

	Experiment	Demonstration	Recountal
1. Analyzing bits of behavior for presence of original nature			
2. Building up learning curves			
3. Mirror drawing			
4. Building up forgetting curves			
5. Testing for individual differences (opposites test, etc.)			
6. Silent reading tests			
7. Arithmetic tests			
8. Experiments on sensation			
9. Waxing and waning of original tendencies			
10. Tests for visual defects			
11. Tests for auditory defects			
12. Trial and error learning			
13. Transfer of training			
14. The work curve to show entrance of fatigue			
15. Illustration of laws of attention:			
(*a*) Effect of intensity			
(*b*) Effect of contrast			
(*c*) Counter-attraction, etc.			
16. Memory; part vs. whole methods			
17. Statistical method:			
(*a*) Central tendencies			
(*b*) Distributions			
(*c*) Correlations			
18. Introspective questionnaire on vividness of imagination			
19. Test for immediate memory such as Knox cube			
20. Apperception:			
(*a*) Interpreting ink blots			
(*b*) Hidden pictures			
(*c*) Word completion			
21. Fluctuation of attention			
22. Ergograph			
23. Experiments on testimony and report			
24. Suggestibility			
25. Normal associations vs. abnormal associations			
26. Eye-movements in reading			
27. Bodily changes in emotion			
28. Pleasantness vs. unpleasantness in color combinations			
29. Testing for IQ			
30. Testing for special abilities			
31. Testing for educational accomplishment			
32. Testing for emotional stability			
33. Associative shifting			
34. Influence of drill			
35. Finding plateaus in learning curves			
36. Questionnaire on children's interests amusements, ambitions, ideals			
37. Puzzles in illustrating trial and error chance variations, multiple response, etc			
38. Motion picture			
39. Survey of play activity			
40. Psychological analysis of misbehavior types			
41. Experiments in animal psychology to illustrate habit formation, situation-connection-response-series			

(*A*) Please list any others not listed above that you have found valuable.

(*B*) Assuming you have reasonable conditions and time for giving your first course in educational psychology please star the experiments which you consider of unquestioned value for relatively immature students.

(*C*) Please list texts and manuals used in your beginning courses.

Name of your institution..

The method of constructing the foregoing letter was that of a rather thorough canvass of all the textbooks and experimental manuals on the market. From these were selected the experiments listed, as

being fairly representative of the general field of experimentation on the subject.

Forty-one returns were obtained—a fair sample, we believe, of the more progressive institutions throughout the country.[1]

---

[1] The names of these institutions follow:

Cornell University, Ithaca, N. Y.
Dartmouth College, Hanover, N. H.
Drake University, Des Moines, Iowa.
Eastern Illinois State Teachers College, Charleston, Ill.
George Peabody College for Teachers, Nashville, Tenn.
Harvard University, Cambridge, Mass.
Illinois State Normal University, Normal, Ill.
Iowa State College, Ames, Iowa.
Iowa State Teachers College, Cedar Falls, Iowa.
Montclair State Normal School, Montclair, N. J.
New York State College for Teachers, Albany, N. Y.
Northeast State Teachers College, Kirksville, Mo.
Oberlin College, Oberlin, Ohio.
Ohio Wesleyan University, Delaware, Ohio.
Pennsylvania State College, State College, Pa.
Purdue University, Lafayette, Ind.
Rutgers College, New Brunswick, N. J.
Sophia Newcomb Memorial College, New Orleans, La.
Southern Branch University of California, Los Angeles, Cal.
State Normal College, Ypsilanti, Mich.
State Normal School, Indiana, Pa.
State Normal School, Lewiston, Idaho.
State Normal School, Fitchburg, Mass.
State Normal School, Stevens Point, Wis.
State University of Montana, Missoula, Mont.
Teachers' College, Kearney, Neb.
Teachers College, New York, N. Y.
Tufts College, Mass.
University of Arkansas, Fayetteville, Ark.
University of California, Berkeley, Cal.
University of Georgia, Athens, Ga.
University of Kansas, Lawrence, Kansas.
University of Maine, Orono, Me.
University of Oklahoma, Norman, Okla.
University of Rochester, Rochester, N. Y.
University of Texas, Austin, Texas.
University of Virginia, University, Va.
University of Washington, Seattle, Wash.
University of Wisconsin, Madison, Wis.
Western State Normal School, Kalamazoo, Mich.
Yale University, New Haven, Conn.

Turning now to the interpretation of the data, let us note a tabulation of the items checked in the respective columns exclusive of "Recountal."

TABLE I[1]

	Experiments	Demonstration	Experiment starred		Experiments	Demonstration	Experiment[2] starred
1	10	12	1	22	2	3	0
2	28	17	18	23	14	8	10
3	18	7	7	24	8	7	0
4	13	10	6	25	6	8	1
5	29	13	15	26	7	9	1
6	20	11	8	27	2	6	1
7	15	12	6	28	9	7	0
8	18	4	5	29	19	19	13
9	0	2	0	30	13	9	8
10	14	18	6	31	15	9	7
11	10	16	5	32	4	4	1
12	20	12	0	33	3	1	1
13	10	7	8	34	10	6	8
14	6	6	1	35	19	7	8
15	17	8	0	36	4	3	1
16	25	10	14	37	8	8	3
17	28	14	11	38	1	2	1
18	21	8	5	39	2	4	3
19	15	10	1	40	1	2	0
20	23	9	7	41	2	5	3
21	14	11	3				

It will be noticed that the following experiments stand out as the most frequently checked both under "Experiments" and "Experiments starred:"

2. Building up learning curves
5. Testing for individual differences
16. Memory—part vs. whole methods
17. Statistical method
29. Testing for IQ. Less, though relatively high importance is given under "Experiments starred" to

[1] The consecutive numbers 1, 2, 3, . . ., 41 correspond to items in the circular letter.

[2] "Experiments starred" refers to answers to question *B* in the circular letter.

13. Transfer of training
23. Experiments on testimony and report
34. Influence of drill
35. Finding plateaus in learning curves.

It may be argued from this that the laboratory part of a beginning course in educational psychology should have these experiments for a nucleus, to which additions might be made as circumstances and available time may allow.

There is, however, a disappointing lack of unanimity of opinion apparent in the foregoing tabulation, both as to what is offered in the laboratory and what is considered of importance for elementary students.

In answer to Question *C* of the letter—a request to list texts and manuals used—the respondents did not in every case give the name of the book, but merely the author. We have therefore thought it best to give only the author's name as representing a general viewpoint of his relative influence, and the corresponding frequency of mention.

Author	Frequency
Angell	2
Averill	3
Bagley	2
Baldwin	1
Betts	2
Breese	1
Bolton	1
Breitweiser	2
Calkins	1
Colvin	7
Dewey	1
Freeman	5
Gordon	2
Halleck	1
Hunter	1
Hollingworth and Poffenberger	1
James	1
Judd	1
Kirkpatrick	2
Langfeld and Allport	1
LaRue	2
Monroe	4
Norseworthy and Witley	2
Ogden	1

Author	Frequency
Parker	1
Peterson, H. A	1
Pillsbury	5
Seashore, C. E	6
Starch	11
Strayer	2
Strong	2
Terman	3
Tracy	1
Thorndike	10
Waddle	1
Watson, John B	1
Woodworth	4

Unless it be argued that an introductory course in general psychology furnishes an adequate psychological background for the teacher, it is obvious that the texts of many of the authors listed hardly fit into a course in educational psychology. They are merely texts adapted for general beginning courses, and the feeling that they are inadequate for the specific purposes of elementary courses in educational psychology is corroborated by the relatively large number of textbooks in educational psychology published within recent years. Ideally, the best way to train the teacher is "on the job." Courses in practice teaching recognize this fact. It is only from consideration of economy and feasibility that all teachers are not so trained. The principle of as direct application of theory to practical problems as possible holds in psychology as well as in teaching problems in general, and the notion that general psychology is sufficient is only one step removed from the much-belabored idea of formal discipline.

The teacher must have not merely a theoretical but primarily a Socratic knowledge of psychology; it must function in the common, every-day come and go of the classroom. There is reason to suspect that too often the prospective teacher's information on matters psychological operates only for purposes of regurgitating for the professor a more or less organized body of abstractions, which are to be relegated—with a sigh of relief—to the mental ash heap immediately after the final examinations. Much the same argument holds concerning several of the experiments listed in the questionnaire and those given in answer to Question *A*.

The following are some of the obviously inept experiments so listed:

Weber's law
Reaction time
Ampullar sense (revolving table)
Threshold differences—skin
Laws of mixture, contrast, and adaptation
Conditions of spatial perception
Study of dreams
Clairvoyance, mediums, automatisms

The thesis here contended for is not that this does not represent useful knowledge, but simply that it is not the most useful knowledge for young teachers in training, who in any case can give but little time to the study of psychology.

The lack of unanimity of opinion concerning what should go into a beginning course in educational psychology is apparent not only in the foregoing, but in the texts that deal directly with educational psychology as well. An attempted analysis of the contents of five of these reveals a decided divergence in point of view, organization, and content; so much so, in fact, that a projected comparison of them here had to be abandoned. The reader may check on this statement by comparing the following:

Starch, D.: Educational Psychology.
Averill, L. A.: Psychology for Normal Schools.
Gordon, Kate: Educational Psychology.
Cameron, E. H.: Psychology and the School.
La Rue, D. W.: Psychology for Teachers.

Making due allowance for the individuality of the author, it must still be true that there is a best way of presenting the subject, as well as a best way of teaching it. What this best way is, yet remains a problem to be solved.

Differences in Viewpoint

That there are some fundamental differences in viewpoint may further be seen by comparing that of Starch and Thorndike, who as indicated by the returns are the leaders in educational psychology. Starch[1] definitely belittles the place of instincts in educational practice. "The direct appeal to, and use of, instinctive reactions in actual concrete instances in school work," he asserts, "are not as frequent and specific as is commonly implied." That this opinion is hardly shared by Thorndike is evidenced by the fact that he devotes the whole of Volume I of his three volume work in Educational Psychology to a discussion of original tendencies.

[1] Starch, D.: "Educational Psychology." 1920, p. 12.

The issue is clearly drawn in a review[1] of Starch's book. "He" (Starch), says the reviewer, "points out that instincts as such have very little significance for education, and that the chief educational doctrines based upon instincts (dynamic theory of instincts, transitoriness, and recapitulation) have very little justification in verified fact. This is a sane view, and we hope that it will tend to neutralize the overemphasis of instincts that has been prevalent in educational discussion since James." It is probably safe to assume that the advertising expert makes his appeal to the same sort of individual, psychologically, as does the educator. Yet the theory of the psychology of advertising is in great part told when instinctive trends have been thoroughly discussed.

Such fundamentally different attitudes as held by Starch and Thorndike cannot both be right. And educational practice cannot be the same whether we hold to the one or the other. If original nature has relatively little weight in the scales, then it follows that a correspondingly larger amount of nurture (*i.e.*, education) will be required to produce the socially efficient individual. If we think of man as a product of original nature and education, the equation:

$$X \times Y = Z$$

would seem to express this relation when X equals original nature, Y equal nurture, and Z equals the end product—man. It is obvious that to have Z remain a constant when either X or Y are changed in value requires a corresponding change in the other member.

Conclusions

1. There is considerable evidence to show that educational psychology is definitely regarded as an experimental science.
2. There is, with a few exceptions, little unanimity of opinion as to what is most important for elementary laboratory courses in the science.
3. Concerning the importance of a few items, as individual differences, statistical method, learning curves, experiments on memory, and testing for intelligence, there is relatively high agreement.
4. A survey of textbooks shows that there is little agreement as to point of view, organization, and content.
5. The two authors most influential in the United States in educational psychology—Starch and Thorndike—disagree fundamentally as to the importance of original nature in the science.

[1] *Jour. Educ. Psy.*, Dec., 1920, pp. 535–6.

A COMPARISON OF MENTAL AGE SCORES OBTAINED BY PERFORMANCE TESTS AND THE STANFORD REVISION OF THE BINET-SIMON SCALE

BUFORD JOHNSON

AND

LOUISE SCHRIEFER

Bureau of Educational Experiments, New York City

The use of non-verbal tests in the determination of age differences in mental traits is recognized as especially important in the early years. The selection of a series of such tests that will differentiate throughout a period of years is still an experimental problem. In the systematic study from year to year of the various age groups in the city and country school we have included the Pintner-Paterson Performance Scale and the Stanford-Revision of the Binet-Simon Scale. This study reports a statistical evaluation of the scores made by the same children when measured by both scales. The children studied were so young both scales could be given during one laboratory examination. No discussion is given of the changes in scores for the individual from year to year. That forms another study which is in progress. All the scores considered were obtained during the year 1920–21. Some of the children had been in the school for several years and had been tested each year. Forty-one children entered the school during this year and had only the one experience with either group of tests. The correlation of each type of scale with chronological age is determined. Scores in other non-verbal tests are also presented and included in the comparative study. These tests do not form the complete series given to these children but from the records we have taken the scores made in the following tests for the various age groups as a basis for this study.

Three-year Group.—Stanford Revision of the Binet; Manikin, Seguin, Mare and Foal of the Pintner-Paterson Performance Scale; the Witmer Cylinder Test; Rossolimo's Pictures No. 1–4.

Four- and Five-year Groups.—Stanford Revision of the Binet; Mare and Foal, Seguin, 5-figure, 2-figure, Casuist, of the Pintner-Paterson Performance Scale; the Witmer Cylinder Test; Action Agent; Rossolimo's Pictures No. 1–7.

Six-year Group.—Stanford Revision of the Binet; Mare and Foal,

Seguin, 5-figure, 2-figure, Casuist, Diagonal, Triangle, Ship, Feature Profile, Knox Cubes of the Pintner-Paterson Performance Scale; the Witmer Cylinder Test; Action Agent: Rossolimo's Pictures No. 1–8.

Seven-year Group.—The six-year tests and in addition Healy *A*; Substitution of the Pintner-Paterson Performance Scale; Rossolimo's Pictures No. 1–10; Dearborn Formboard.

Eight- and Nine-year Groups.—Stanford Revision of the Binet: Complete Pintner-Paterson Performance Scale with the exception of the Adaptation Board; Healy Picture Completion II.

The Method.—The directions for giving and for scoring these tests may be found in the following publications: Terman: "The Measurement of Intelligence," Houghton Mifflin; Pintner-Paterson: "A Scale of Performance Tests," Appleton; Bureau of Educational Experiments: "Health Education and the Nutrition Class," E. P. Dutton.

The group of 86 children is made up of both boys and girls ranging in age from 3 to 9.8 years. The age groups are formed on the basis of the age in years that each child had reached at the time of the testing. Children in the 3-year group had reached their third birthday but had not yet reached their fourth; consequently the theoretical average for each age group falls at 3.5, 4.5 years, etc.

Although the Jewish and Italian nationalities are represented, it is primarily an American born group, its members coming from families widely variant as to fields of activities, professional and industrial. Including mothers and fathers, we find 90 parents are represented. The number of children in a family range from one to four. The classification according to vocation shows that 36 per cent are in the professions; 17 per cent are artists; 37 per cent are in commercial occupations, and 10 per cent are in miscellaneous activities, such as teamster and detective. It seems a representative group of city children, many of whom were born outside of New York City. Admission to the school is not wholly dependent upon ability to pay the tuition fees as there are some scholarships available. Other factors that enter into the selection of the group, aside from the exclusion of children below the normal rating in intelligence, are the possibilities of continued attendance of the same children throughout a period of years, and the cooperation of the parents in an educational scheme involving a research program where such cooperation is especially desirable.

The Stanford Revision IQ's range from 96–167, with an average of

114.3 s.d. 11.3. This average places the group in the superior intelligence class as rated by Terman. The distribution is as follows.

IQ	Per Cent	
96–105	17 or	20
106–115	37 or	43
116–125	18 or	21
126–135	12 or	14
136–145	1 or	1
146–155	—	
156–165	—	
166–175	1 or	1
	86	100

In the Terman study of 905 children between 5–14 years of age, 69 per cent of the children obtained an IQ of 96 or more; in the City and Country School 100 per cent of the children obtained an IQ of 96 or more, 25 per cent being in the very superior intelligence class according to the Terman rating, that is, IQ 120 or more. We do not believe this is a very unusual distribution for children of these ages in private schools. The young child in a stimulating environment has an intelligence quotient as determined by the Stanford Revision that we do not expect to be constant but to decrease with increasing years. Slight variations in scores make greater differences in intelligence quotients than are possible in later years.

The Pintner-Paterson Tests were not standardized for the lower ages so that it was necessary for us to determine median age limits from our own test results in order that a rating might be made of the younger children. According to the method described in the Pintner-Paterson Scale (p. 151) for determining the limiting values for various ages we have determined 3-year and 4-year limits for Seguin and Mare and Foal, using the available Pintner-Paterson norms and modifying our own where over-lapping occurred; that is, Pintner-Paterson have set their lower 5-year limit in time for Seguin Formboard at 50 seconds. Our lower limit for 4-years would fall at 44 seconds, obviously conflicting with the Pintner-Paterson 5-year norm. We have, therefore, modified our 4-year limits and used the following:

	3-year	4-year	5-year
Seguin time in seconds..........	110– 55	54– 51	50–32
	300–226	225–151	150–89

These test values have been determined for only a few cases including eight 3-year-old children and fifteen 4-year-olds.

The following table contains the averages of chronological age, of Stanford Revision Mental Age, and the Pintner-Paterson Median Age for each of the year groups:

Age group	Number of cases	Chronological age in months		Terman mental age in months		Pintner-Paterson median age in months	
		Average	σ	Average	σ	Average	σ
3	8	40.5	4.27	48.0	6.63	58.5	13.99
4	15	53.5	3.75	60.4	5.76	64.8	17.83
5	21	66.3	3.32	76.9	8.52	89.9	25.42
6	21	75.7	2.90	86.7	6.07	109.6	19.64
7	7	89.3	2.62	104.4	10.78	143.1	29.13
8	5	102.4	3.38	108.4	4.96	134.4	19.20
9	9	115.0	2.83	129.8	12.57	146.0	17.89

Computing the differences between the Stanford Revision and Pintner-Paterson measures for each year group and relating them to the probable error of their differences, we find a significant difference for each age except the 4-year, which shows only a possible difference. The largest difference occurs in the 6-year group, the next largest at 7 and 8, which again suggests the greater usefulness of the Performance scale at the ages of 6, 7, and 8 years.

PROBABLE ERRORS OF DIFFERENCE BETWEEN PINTNER-PATERSON AND TERMAN, RELATED TO THE ACTUAL DIFFERENCE

AGE GROUP	$\frac{D}{PED}$
3	$\frac{10.5}{3.68} = 2.85$
4	$\frac{4.4}{3.25} = 1.35$
5	$\frac{13.00}{3.94} = 3.29$
6	$\frac{22.9}{3.02} = 7.58$
7	$\frac{38.7}{7.92} = 4.88$
8	$\frac{26.0}{5.99} = 4.34$
9	$\frac{16.2}{4.91} = 3.30$

There are 22 children having a Stanford Revision IQ of 120 or more. These would be considered in the very superior group of intelligence. They distribute themselves as follows:

Age.......	III	IV	V	VI	VII	VIII	IX
	3	2	6	6	2	0	3

and comprise one-fourth of the entire group. The distribution shows that approximately one-third of each of the separate age groups is included, except in the Four Group, where only 2 out of 15 children have very superior intelligence according to the Stanford Revision rating, and in the Eight Group, where there are none.

Comparing these Stanford Revision IQ with IQ's similarly calculated for the Pintner-Paterson Median Age, we find that 19 of the 22 have Pintner-Paterson IQ's of 150 or more. The remaining 3 IQ's are 140, 133, 133, which in ordinary selection would be considered a high rating. This similarity in scores suggests that children of high general intelligence as measured by the Stanford Revision scale are also high in performance tests.

Upon the suggestion of Pintner and Paterson (p. 157) that the children above the middle 50 per cent of a given troup are probably bright, we have selected from each of our age groups the children of the upper 25 per cent in Pintner-Paterson Rating; this forms a basis for comparison with the very superior group according to the Stanford Revision Rating. The following table shows the distribution of this grouping; obviously. it includes approximately one-fourth of the entire group.

Pintner-Paterson Median Age Distribution for Upper 25 Per Cent of Each Year Group

Year group	Median age								
	6	7	8	9	10	11	12	13	14
3	1	1							
4	..	2	2						
5	..	..	..	2	1	..	2		
6	..	..	..	..	..	3	2		
7	..	..	..	..	..	..	..	..	2
8	..	..	..	..	..	..	..	..	1
9	..	..	..	..	..	..	..	..	2

Comparing these ratings with the Stanford Revision ratings of the same children, we do not find the agreement that was noted above. Only eight of these children would be included in the very superior

intelligence class according to the Stanford Revision. The remaining 13 are distributed as follows:

Stanford Revision IQ....	95–99	100–104	105–109	110–114	115–119
	2	..	5	4	2

The two children below 100 in Stanford Revision IQ are in the 7- and 9-year groups. Analyzing their performance in individual tests, we find the 7-year old considerably above his class median in 6 out of 8 tests; in 2-figure time the class median was 45, the individual's time 65 seconds; in Knox Cubes the median score was 6, the individual's score 5. In the Stanford Revision he succeeded in the reproduction of familiar things, such as counting backwards, the date, easy definitions, and in weight discrimination. He failed, however, in tests requiring language expression and abstractions, *i.e.*, comprehension, similarities, sentences, rhymes, vocabulary. The 9-year boy falls below his group median in only one test of 14; in 8 tests he is above the median. Even in the Knox Cube test, which is primarily a memory test, and in Picture Completion and Substitution, which involve

Table I.—Median Age for Individual Test Scores

Year group	Mare and Foal time	Seguin time	5-figure time	5-figure errors	2-figure time	2-figure moves	Casuist time	Casuist errors	Triangle time	Triangle errors	Diagonal time	Diagonal errors	Healy A time	Healy A moves	Manikin score	Feature time	Ship score	Picture completion	Substitution	Cube test	Group median age
3	5	3	..	....	..	....	..	..	..	..	..	..	..	..	6	..	....	....	....	..	5.0
4	6	5	6	6.0	6	7.0	6	..	..	..	..	..	..	..	7	..	....	....	....	..	6.0
5	7	5	7	10.0	7	8.0	7	10	10	15	6	13	..	..	7	..	5.0	....	....	7	7.0
6	..	6	8	12.5	8	9.0	8	12	7	15	7	13	..	..	..	15	7.0	....	....	7	8.0
7	..	..	13	15.0	10	12.5	12	15	9	15	12	14	7	12	..	..	12.0	....	....	9	12.0
8	10	10	12	15.0	10	10.0	13	12	11	15	14	14	7	15	8	12	9.5	9	10	12	11.5
9	13	12	14	15.0	11	10.0	10	12	13	15	11	12	7	15	8	15	12.0	11	10	14	12

complex mental processes, the child runs ahead of his class median. In the Stanford Revision his performance was similar to that of the 7-year old, in that he passed tests involving reproduction and memory such as word naming in the 10-year old series, and the repetition of digits backward in the 12-year series; he gave a fairly creditable performance in Comprehension of Questions. The tendency is to pass tests involving construction from concrete màterials but to fail in tests emphasizing comprehension and abstract reasoning as the Stanford Revision tests do.

These results again suggest that superior general intelligence is accompanied by a high degree of ability in performance. In medium levels this is not consistently true. A high performance rating may be made by one who shows mediocre ability in general intelligence tests.

In Table I we have tabulated the median ages which correspond to the median test scores obtained by the children of our group. These individual medians have been summarized in the last column under "Group Median Age." It is evident that each of the year groups, 3, 4, 5, 6, are 2 years ahead of the Pintner-Paterson standards, the 3-year group having a 5-year rating, the 4-year group a 6-year rating, etc. The 7-year group has a median age of 12, which is undoubtedly too high to be accurate, but the interesting fact is that the scale fails to differentiate between the 7-, 8-, and 9-year group. Some differentiation of course would be expected since 7 tests have 14-year old norms and 3 tests have 15-year-old norms.

Continuing the method referred to above of determining median age limits, in Table II we present these limits for our group. They are only approximations, owing to an arbitrary determination of the limits where intermediate ages were missing. The lower limits in such cases have been determined by the lower quartile of a given age, the upper limits by the upper quartile. They are open to criticism and are presented merely for their suggestive value. Substitution, Healy A, Picture Completion and Knox Cubes have been omitted

Table II.—Median Intervals for Scoring Tests
Determined by results obtained in the city and country school

Year group	3	4	5	6	7	8	9
Mare and Foal	170–91	90–66	65–50	......	.....	49–33	32–0
Seguin	110–55	54–44	43–33	32–31	.....	17–15	14–0
Manikin	3	4	4	......	.....	5	5
5-figure time	......	300–181	180–125	124–79	78–53	52–43	42–0
2-figure time	......	300–146	145–97	96–61	60–45	44–43	42–0
Casuist time	......	30–238	237–146	145–85	84–64	.....	63–0
Triangle time	......	....	....	120–67	66–54	53–40	39–0
Diagonal time	......	....	300–98	97–56	55–25	24– 0	
Ship score	......	....	6–12	13–18	19–20		

because only a very limited number of measures were available; errors and moves in 5-figure, 2-figure, Casuist, Triangle, and Diagonal showed very little differentiation.

TABLE III

		Chronological age		Terman mental age		Pintner-Paterson median age	
		Number of cases	Correlation coefficient	Number of cases	Correlation coefficient	Number of cases	Correlation coefficient
Mare and Foal......	Time	53	0.554	53	0.584	53	0.582
Seguin.............	Time	63	0.650	63	0.688	63	0.504
Five figure board...	Time	65	0.586	65	0.603	65	0.697
Five figure board...	Errors	..	....	61	0.401		
Two figure board...	Time	60	0.579	60	0.335	60	0.441
Two figure board...	Errors	..	....	58	0.200		
Casuist............	Time	64	0.503	64	0.536	64	0.516
Casuist............	Errors	..	....	63	0.203		
Triangle...........	Time	60	0.298	60	0.335	60	0.387
Triangle...........	Errors	..	....	58	0.120		
Diagonal...........	Time	51	0.212	51	0.289	51	0.592
Diagonal...........	Errors	..	....	50	0.217		
Healy Puzzle *A*.....	Time	16	0.040	16	−0.205	16	0.131
Healy Puzzle *A*.....	Moves	..	....	16	0.249		
Manikin...........	Score	52	0.533	52	0.569	52	0.657
Feature profile......	Time	25	0.270	25	0.267	25	0.385
Ship...............	Score	57	0.534	57	0.952	57	0.632
Picture completion..	Score	13	0.548	13	0.243	13	0.532
Substitution........	Score	13	0.174	13	0.230	13	0.488
Knox cubes.........	Score	56	0.517	56	0.536	56	0.546
Cylinders..........	Time	86	0.565	86	0.576	86	0.669
Rossolimo..........	Time	80	0.264	80	0.504	80	0.340
Rossolimo..........	Score	83	0.812	83	0.836	83	0.853
Action agent........	Time	62	0.360	62	0.367	62	0.292
Action agent........	Score	62	0.455	62	0.473	62	0.349
Healy picture completion II........	Score	13	−0.016	13	0.204	13	−0.147
Dearborn II........	Time	19	0.218	19	0.400	19	0.354
	Moves	..	....	19	0.083		
Chronological age...	....	..	....	86	0.946	86	0.782
Terman mental age.	....	..	....	..	....	86	0.818

The curves for time measures suggest that time is a good measure up to a certain point in a child's development; having attained a certain limit, however, it is rarely possible for an individual as he increases his age to lessen the time of a given performance. If he can lessen it, practice and exceptional motivation have probably made it a special ability.

The correlations with age show more plainly this relationship between age and a given test. The following correlations are arranged in rank order:

Chronological age with	Seguin	0.650
	5-figure time	0.586
	2-figure time	0.579
	Mare and Foal time	0.554
	Picture completion score	0.548
	Ship score	0.534 (limited number)
	Manikin score	0.533
	Knox score	0.517
	Casuist time	0.503

There is a sharp line of demarcation between the above mentioned tests and the remaining ones that arranged in rank order are:

Chronological age with	Triangle time	0.298
	Profile time	0.270
	Diagonal time	0.212
	Substitution score	0.174
	Healy a time	0.040

Discussion and Summary

Judging from fineness of discrimination, age progression and correlation, the following tests prove themselves most adequate; Mare and Foal, Seguin, 5-figure, 2-figure, Casuist, Manikin, Ship, and Cubes. Of the remaining tests, Triangle and Diagonal are fairly good. Picture Completion and Substitution are untried because of the limited number of measures, but the indications are that they would fall in the desirable group. These tests correspond closely to the Pintner-Paterson Short Scale which includes Mare and Foal, Seguin, 5-figure, 2-figure, Casuist, Manikin, Feature and Ship, Picture Completion and Cubes. We would omit, however, from the Performance Scale "errors" and "moves" as measures on the basis of lack of age progression and differentiation between the ages, also difficulty of recording.

In addition to the Stanford Revision and Pintner-Paterson Ratings,

we obtained a score in the Dearborn Group tests of Intelligence, Series I. The average score in Dearborn was 73, this being the norm for 10½-year old children. Our children average 9.2 years at time of testing.

Dearborn correlated 0.55 with Stanford Revision Mental Age and 0.83 with Pintner-Paterson Median Age. The high correlation between Dearborn and Pintner-Paterson would be expected from the fact that language expression plays no part in either of them, the chief requirement being to execute directions quickly and accurately. It would be interesting to follow up the relationship between the two measures, because of the time saving involved in giving the Dearborn Test instead of the Pintner-Paterson Scale. The correlation between Dearborn and Stanford Revision is not high.

The Stanford Revision Scale correlates more highly with Chronological Age than with the Performance Scale, but correlates more highly with the Performance Scale than does the Performance Scale with Chronological Age.

Specific tests in the Performance Scale have fairly high correlation-coefficients with Chronological Age and with the Mental Age scores obtained by both scales.

Those who made high intelligence quotients by the Stanford Revision make high ratings in Performance Tests. There are some who make high Performance Scale ratings but hardly attain average Intelligence Quotients.

The Rossolimo series of graded pictures is the best single performance test as measured by high correlation with the Stanford Revision Scale.

It is indicated that a series of Performance Tests may be selected which will give ratings comparable with the Intelligence Quotients and will show good age progression. This series is especially desirable for use with young children in order that we may supplement data which are obtained by methods that put a premium upon vocabulary attainments.

CORRELATIONS OF FOUR INTELLIGENCE TESTS WITH GRADES

A. M. JORDAN

University of Arkansas

There are two ways of progressing in the development of mental tests: The first is to invent new tests, while the second is to refine and improve those extant. In order to develop or refine a test, knowledge of its advantages and limitations is necessary. This knowledge is gained by correlating the test as a whole, and the several elements of the test with various criteria. From the criteria which might be used the one chosen in this study consists of the grades made by the students in their subjects of instruction. The reliability coefficient of grades in successive semesters is around 0.75 so that the criterion chosen varies.

The purpose of the investigation is: First, to find out the group test of the elements of the group test which furnish the best prognosis of the standing of pupils in the high school subjects of instruction when the total grades are combined into one score. The second purpose is to discover the test which correlates most highly with the grades received in English, mathematics, general science and history.

The method employed was: First, to give four intelligence tests (Army Alpha, Terman, Otis, and Miller) to 67 high school pupils;[1] second, to give the Army Alpha to 315 university students. With this material collected the next task was to proceed with the correlations. Correlations together with probable errors were computed between (*a*) each of the four group tests given to high school pupils and the total points of the subjects of instruction. The value of each subject was determined by grade points: $A = 6$, $A-$ and $B+ = 5$, $B = 4$, $B-$ and $C+ = 3$, $C = 2$, $C-$ and $D+ = 1$, $D = 0$, $E = -1$, and $F = -2$. (*b*) Each of the tests was correlated with English, mathematics, general science, and history individually, while (*c*) each of the elements of each test (31 in all) was correlated with all subjects combined, and with English, mathematics, general science and history individually. Correlations were also made between Army Alpha and university subjects of instruction for the first, second and third terms of 1 year.

[1] These four tests were given by S. R. Powers of the University of Minnesota to the pupils in the University of Arkansas Training High School in 1921.

In Tables I to X are the coefficients of correlation with their probable errors. In all 215 correlations were made in order to find out which group test or individual test correlated most highly with the subjects of instruction.

Of these intelligence tests Otis, Terman, and Army Alpha are familiar. The Miller test was developed by W. S. Miller of the University of Minnesota and is called "Test for High School Pupils." It was standardized by the administrative section of the high school conference. There are three parts to the test. The first consists of 40 disarranged sentences and requires 8 minutes to give. The second deals with causes and effects—*e.g., fire* (hot house damage) "damage" being underlined correctly. There are 40 of these and 5 minutes are consumed in giving it. The third is the well known analogies and 6 minutes are given to answer the 40 analogies.

Table I.—Correlations of Four Group Tests of Intelligence with the Combined Grade Points of All Subjects (67 Cases)

	Fall	Winter	Spring	Average
Otis	0.414 ± 0.066	0.527 ± 0.058	0.409 ± 0.067	0.450
Army Alpha	0.458 ± 0.064	0.449 ± 0.060	0.521 ± 0.060	0.476
Miller	0.460 ± 0.063	0.511 ± 0.059	0.457 ± 0.060	0.476
Terman	0.483 ± 0.062	0.476 ± 0.063	0.517 ± 0.059	0.492

Table I sets forth the correlations obtained for successive terms between each of the four group tests and the total number of grade points received throughout the year by the high school pupils in their several subjects of instruction. It will be noted that the Terman test stands highest in the fall quarter, Miller in the winter, and Army in

Table II.—Correlations between Each Element of Each of the Group Tests of Intelligence with the Combined Grade Points of All Subjects (67 Cases)

Otis	Subjects	Army	Subjects	Miller	Subjects	Terman	Subjects
1	0.269 ± 0.076	1	0.423 ± 0.067	1	0.466 ± 0.064	1	0.555 ± 0.057
2	0.342 ± 0.072	2	0.460 ± 0.064	2	0.380 ± 0.070	2	0.267 ± 0.076
3	0.305 ± 0.075	3	0.447 ± 0.065	3	0.441 ± 0.066	3	0.382 ± 0.070
4	0.323 ± 0.073	4	0.391 ± 0.069	..	....	4	0.191 ± 0.079
5	0.479 ± 0.063	5	0.514 ± 0.059	..	....	5	0.443 ± 0.066
6	0.374 ± 0.071	6	0.371 ± 0.070	..	....	6	0.461 ± 0.063
7	0.414 ± 0.068	7	0.413 ± 0.068	..	....	7	0.492 ± 0.062
8	0.293 ± 0.075	8	0.244 ± 0.077	..	....	8	0.444 ± 0.066
9	0.298 ± 0.075	..	....	..	....	9	0.403 ± 0.069
10	0.136 ± 0.080	..	....	..	....	10	0.367 ± 0.071

the spring. Possibly the most significant finding here is the slight difference exhibited among the size of the correlations, no correlation being below 0.4 and no one being as high as 0.53. Moreover, no correlations are high, the results agreeing well with the 0.4 to 0.5 usually found by other investigators. Terman's tests average a little but not much above the others.

More interesting for our purposes is Table II which contains the correlations of each element of each group test with the total grade points of all subjects. Thus there are 31 correlation coefficients which vary in size from 0.136 for Otis-10, a test of memory, to 0.555 for Terman-1, a test of general information.

The highest correlations between individual tests and the total grade points of all subjects for the entire year were as follows:

Terman-1	General informaion	0.555 ± 0.057
Army -5	Mixed sentences	0.514 ± 0.059
Terman-7	Analogies	0.492 ± 0.062
Otis-5	Arithmetic problems	0.479 ± 0.063
Miller-1	Mixed sentences	0.466 ± 0.066

These correlations while not especially high are significant. It is to be noted that "mixed sentences" occurs twice. When these correlations are compared with those obtained with the four group tests and the total grades it is found that two of the individual tests mentioned in Tables I and II have higher correlations with all the grades than has the sum of the individual tests combined into the group of tests commonly known as the Otis tests of intelligence, Terman, etc. For example, the Otis test has a coefficient of 0.450; the Army test of 0.476; the Miller test of 0.476; and the Terman test of 0.492. It is evident, therefore, that Terman-1 (general information) which requires only 2 minutes to give and 1½ minutes to score would place pupils in groups, graded according to their capacities to learn as measured by the grades received, more correctly than would any one of the four group tests. Compare 2 minutes with an hour in giving, and 1½ minutes with 8 or 10 in scoring and the time for practical purposes seems clear.

The criticism might be raised, however, that a group of individuals might react differently when the test was given as a separate entity and for only a short period of time than they would when the test was one of a group extending over a longer period of time. I have no experimental evidence either for or against this problem. If this condition were found to be true it would not vitiate the use of this test

as one element of a group for prognostic purposes, nor prevent time being saved in scoring.

The correlations with English are seen in Table III. Here are found both the correlations obtained between the group tests as a whole and grades in English and the correlations between individual elements of each of the four group tests and English. The size of the correlations varies from Terman-2 with 0.14 to Terman-3 with 0.57.

TABLE III.—CORRELATIONS BETWEEN EACH ELEMENT OF EACH OF THE FOUR GROUP TESTS OF INTELLIGENCE AND GRADES IN ENGLISH (64 CASES)

Otis	English	Army	English	Miller	English	Terman	English
1	0.237 ± 0.077	1	0.301 ± 0.074	1	0.594 ± 0.053	1	0.433 ± 0.066
2	0.340 ± 0.071	2	0.386 ± 0.082	2	0.433 ± 0.066	2	0.141 ± 0.080
3	0.399 ± 0.069	3	0.544 ± 0.057	3	0.393 ± 0.069	3	0.572 ± 0.055
4	0.421 ± 0.067	4	0.527 ± 0.059	..	....	4	0.386 ± 0.069
5	0.425 ± 0.067	5	0.409 ± 0.068	..	....	5	0.253 ± 0.076
6	0.456 ± 0.065	6	0.279 ± 0.075	..	....	6	0.492 ± 0.063
7	0.313 ± 0.073	7	0.301 ± 0.074	..	....	7	0.552 ± 0.057
8	0.225 ± 0.078	8	0.388 ± 0.069	..	....	8	0.459 ± 0.064
9	0.189 ± 0.074	..	....	..	....	9	0.383 ± 0.069
10	0.212 ± 0.078	..	....	..	....	10	0.387 ± 0.069
Total test....	0.466 ± 0.065	..	0.472 ± 0.065	..	0.564 ± 0.057	..	0.523 ± 0.061

English is one of the subjects with which a considerable number of individual tests correlate significantly, there being no fewer than 10 that have a correlation of 0.4 or above. The number of pupils was 64. These correlations are:

Miller-1	Mixed sentences..................................	0.594 ± 0.053
Terman-3	Opposites..	0.572 ± 0.055
Terman-7	Analogies..	0.552 ± 0.057
Army-3	Reasons..	0.544 ± 0.057
Army-4	Opposites..	0.527 ± 0.059
Terman-6	Sentence meaning...............................	0.492 ± 0.063
Terman-8	Mixed sentences..................................	0.459 ± 0.064
Otis-6	Geometrical figures.............................	0.456 ± 0.065
Otis-5	Arithmetic problems............................	0.425 ± 0.067
Otis-4	Proverbs..	0.421 ± 0.067

These correlations between individual tests and English may be compared with the correlations with English when the group tests are used.

Otis..	0.466 ± 0.065
Army......................................	0.472 ± 0.065
Miller.....................................	0.564 ± 0.057
Terman....................................	0.233 ± 0.061

One individual test, Miller-1 (mixed sentences), has a higher coefficient of correlation than the highest group test. It takes 8 minutes to give Miller-1 while it takes 20 minutes to give the Miller intelligence test; yet Miller-1 correlates with English more highly (0.594) than does the Miller intelligence test (0.564).

Again comparison may well be made between the correlations obtained for the whole year with those obtained for each term when the group tests were used. It is evident from Table III and Table IV that considered as a whole the correlations are somewhat smaller with term grades than with year grades. This is to be expected since, in general, to raise a correlation in the most convenient way is to lengthen the test.

TABLE IV.—CORRELATIONS BETWEEN FOUR GROUP TESTS OF INTELLIGENCE AND ENGLISH FOR SUCCESSIVE TERMS (67 CASES)

	Fall	Winter	Spring	Average
Otis	0.390 ± 0.067	0.405 ± 0.068	0.443 ± 0.067	0.412
Army	0.411 ± 0.067	0.401 ± 0.068	0.361 ± 0.072	0.391
Miller	0.426 ± 0.066	0.430 ± 0.067	0.430 ± 0.067	0.428
Terman	0.403 ± 0.067	0.364 ± 0.071	0.507 ± 0.062	0.424

The individual tests (Table V) which correlate most highly with the grades in mathematics are for the most part those concerned with mathematical operations. The one exception is the comparatively high correlation of hard oral directions. There are 47 cases.

TABLE V.—CORRELATION OBTAINED BETWEEN EACH ELEMENT OF THE FOUR GROUP TEST OF INTELLIGENCE AND GRADES IN MATHEMATICS (47 CASES)

Otis	Mathematics	Army	Mathematics	Miller	Mathematics	Terman	Mathematics
1	0.294 ± 0.090	1	0.608 ± 0.062	1	0.405 ± 0.082	1	0.316 ± 0.088
2	0.252 ± 0.092	2	0.521 ± 0.072	2	0.464 ± 0.077	2	0.207 ± 0.094
3	0.035 ± 0.098	3	0.406 ± 0.082	3	0.439 ± 0.073	3	0.291 ± 0.090
4	0.283 ± 0.090	4	0.233 ± 0.093	..	....	4	0.074 ± 0.098
5	0.676 ± 0.053	5	0.212 ± 0.094	..	....	5	0.529 ± 0.071
6	0.526 ± 0.071	6	0.471 ± 0.076	..	....	6	0.249 ± 0.092
7	0.357 ± 0.086	7	0.454 ± 0.077	..	....	7	0.342 ± 0.088
8	0.278 ± 0.090	8	0.274 ± 0.091	..	....	8	0.300 ± 0.093
9	0.222 ± 0.093	..	....	..	....	9	0.202 ± 0.130
10	0.201 ± 0.094	..	....	..	....	10	0.448 ± 0.078
Group test...	0.430 ± 0.079	..	0.511 ± 0.073	..	0.456 ± 0.077	..	0.436 ± 0.079

The highest correlations are:

Otis-5	Arithmetic problems	0.676 ± 0.053
Army-1	Oral directions	0.608 ± 0.062
Terman-5	Arithmetic problems	0.529 ± 0.071
Otis-6	Geometric figures	0.526 ± 0.071
Army-2	Arithmetic problems	0.521 ± 0.071

It is to be noted that three of these are arithmetical problems and one is concerned with geometrical figures. Contrast these correlations with those obtained from the group tests.

Army	0.511 ± 0.073
Miller	0.456 ± 0.077
Otis	0.430 ± 0.079
Terman	0.436 ± 0.079

Therefore, if there are large classes in mathematics and divisions of them must be made, approximately homogeneous groups may be

TABLE VI.—CORRELATIONS BETWEEN FOUR GROUP TESTS OF INTELLIGENCE AND MATHEMATICS FOR SUCCESSIVE TERMS (47 CASES)

	FALL	WINTER	SPRING	AVERAGE
Otis	0.300 ± 0.080	0.374 ± 0.080	0.170 ± 0.090	0.281
Army	0.302 ± 0.079	0.366 ± 0.080	0.444 ± 0.085	0.371
Miller	0.299 ± 0.080	0.389 ± 0.080	0.338 ± 0.094	0.342
Terman	0.296 ± 0.080	0.287 ± 0.086	0.423 ± 0.086	0.335

obtained in a short period of time by using Otis-5. Furthermore, comparisons between the correlations obtained for the whole year in

TABLE VII.—CORRELATIONS OBTAINED BETWEEN EACH ELEMENT OF EACH OF THE FOUR GROUP TESTS OF INTELLIGENCE AND GRADES IN GENERAL SCIENCE (32 CASES)

Otis	General science	Army	General science	Miller	General science	Terman	General science
1	0.306 ± 0.108	1	0.491 ± 0.090	1	0.567 ± 0.098	1	0.502 ± 0.089
2	0.513 ± 0.088	2	0.513 ± 0.088	2	0.513 ± 0.086	2	0.323 ± 0.106
3	0.280 ± 0.110	3	0.468 ± 0.093	3	0.423 ± 0.080	3	0.468 ± 0.093
4	0.451 ± 0.095	4	0.514 ± 0.088	..	....	4	0.402 ± 0.100
5	0.524 ± 0.086	5	0.548 ± 0.083	..	....	5	0.381 ± 0.102
6	0.429 ± 0.097	6	0.245 ± 0.112	..	....	6	0.516 ± 0.087
7	0.278 ± 0.110	7	0.392 ± 0.100	..	....	7	0.546 ± 0.084
8	0.104 ± 0.118	8	0.427 ± 0.097	..	....	8	0.413 ± 0.098
9	0.287 ± 0.109	..	....	..	....	9	0.561 ± 0.082
10	0.345 ± 0.105	..	....	..	....	10	0.369 ± 0.103
Group test....	0.502 ± 0.89	..	0.596 ± 0.77	..	0.592 ± 0.78	..	0.636 ± 0.071

mathematics (Table V) and for those obtained for each term (Table VI) show smaller correlations for each term than for the whole year. In all cases save one all of the correlations for the whole year are above any one for the term.

General science considered from the standpoint of mental processes involved seems to be truly of a general nature and to resemble in its demand processes similar to those demanded of the individual tests (Table VII) since there are eleven of these tests with correlations higher than 0.5 with this subject. The number of cases was 32. These eleven highest correlations are:

Miller-1	Mixed sentences	0.567 ± 0.080
Terman-9	Classification	0.561 ± 0.080
Miller-5	Mixed sentences	0.548 ± 0.083
Terman-7	Analogies	0.546 ± 0.084
Miller-2	Giving effect of words	0.531 ± 0.086
Otis-5	Arithmetic problems	0.524 ± 0.086
Terman-6	Sentence meaning	0.516 ± 0.087
Army-4	Opposites	0.514 ± 0.088
Army-2	Arithmetic problems	0.513 ± 0.088
Otis-2	Opposites	0.513 ± 0.088
Terman-1	Information	0.502 ± 0.089

The group tests also correlate significantly with general science.

Otis	0.502 ± 0.089
Army	0.596 ± 0.077
Miller	0.592 ± 0.078
Terman	0.636 ± 0.071

In the case of general science three out of four of the group tests correlate more highly with general science than does any single one of the individual tests. Particularly noteworthy is the correlation of 0.636 in the case of the Terman test. Tables VII and VIII indicate as in the preceding subjects a smaller correlation for the term than for the three terms combined into 1 year.

Table VIII.—Correlations between Four Group Tests of Intelligence and General Science for Successive Terms

	Fall	Winter	Spring	Average
Otis	0.432 ± 0.096	0.333 ± 0.103	0.456 ± 0.094	0.407
Army	0.539 ± 0.084	0.396 ± 0.098	0.506 ± 0.074	0.480
Miller	0.476 ± 0.092	0.388 ± 0.099	0.536 ± 0.080	0.466
Terman	0.485 ± 0.091	0.453 ± 0.094	0.586 ± 0.077	0.508

Since there are only 20 cases in history (Table IX) the results are not reliable. This unreliability is reflected in the unusually large probable errors.

Table IX.—Correlations Obtained between Each Element of Each of Four Group Tests of Intelligence with the Grades in History (20 Cases)

Otis	History	Army	History	Miller	History	Terman	History
1	−0.213 ± 0.149	1	0.253 ± 0.141	1	0.288 ± 0.138	1	0.362 ± 0.131
2	0.083 ± 0.149	2	0.115 ± 0.148	2	−0.047 ± 0.150	2	0.168 ± 0.146
3	0.193 ± 0.145	3	0.471 ± 0.117	3	0.033 ± 0.150	3	0.365 ± 0.127
4	0.247 ± 0.141	4	0.136 ± 0.147	..	....	4	0.011 ± 0.151
5	0.079 ± 0.150	5	0.421 ± 0.124	..	....	5	0.397 ± 0.127
6	0.001 ± 0.150	6	0.196 ± 0.145	..	....	6	0.588 ± 0.098
7	0.199 ± 0.147	7	0.007 ± 0.151	..	....	7	0.364 ± 0.131
8	0.135 ± 0.148	8	0.237 ± 0.142	..	....	8	0.364 ± 0.131
9	0.392 ± 0.127	..	....	..	....	9	0.429 ± 0.121
10	−0.008 ± 0.150	..	....	..	....	10	0.117 ± 0.149
Group test....	0.262 ± 0.140	..	0.319 ± 0.136	..	0.168 ± 0.148	..	0.408 ± 0.121

The following individual tests correlated highest with history:

Terman-6	Sentence meaning..............................	0.588 ± 0.098
Army-3	Selecting reasons..............................	0.471 ± 0.117
Terman-9	Classification..............................	0.429 ± 0.121
Army-5	Mixed sentences..............................	0.421 ± 0.124

The ability to get meaning from a series of sentences some of which are false has the highest correlation with history. The group tests which in no case correlate so highly with history have the following coefficients:

Army....................................	0.319 ± 0.136
Terman....................................	0.408 ± 0.121
Otis....................................	0.262 ± 0.140
Miller....................................	0.168 ± 0.148

By reference to Table X it is once more evident that the correlations for 1 year are higher in general than those for individual terms.

Table X.—Correlations between Four Group Tests of Intelligence and History for Successive Terms (20 Cases)

	Fall	Winter	Spring	Average
Otis...........	0.248 ± 0.146	0.319 ± 0.120	0.453 ± 0.114	0.340
Army..........	0.341 ± 0.130	0.256 ± 0.120	0.258 ± 0.132	0.285
Miller..........	0.198 ± 0.141	0.101 ± 0.135	0.141 ± 0.132	0.147
Terman........	0.419 ± 0.124	0.313 ± 0.125	0.335 ± 0.122	0.356

The correlations of Army Alpha, Otis and Miller are unreliable because they are not three times as large as their respective probable

errors, or, if the more stringent test of the coefficient's being four times its probable error should be applied, not one of the group tests and only Terman-6 (sentence meaning) and Army-3 (selecting reasons) are sufficiently reliable. Thus again the advantages of the individual tests are apparent.

In general except in the case of general science it is not at all difficult to find individual tests which correlate more highly with any or all subjects of instruction than the widely used group tests. When the value of each element of each test is accurately determined the question of the purpose to which a test is to be put will, of necessity, be an important item in its selection. The particular question raised here is one of prognosis, the human engineer wishing to know how prophetic the standing in a mental test is of the standing in some life situation.

In this paper no attempt has been made to compare the findings with those obtained in other studies. This, if complete, would mean a paper two or three times as long as the present one. However, the correlations obtained with the Army Alpha and grades of university freshmen might be set forth for comparative purposes:[1] From Table XI it may be seen that, in general the coefficients of correlation are somewhat but not much higher when the high school grades are used than when the university grades are used.

TABLE XI.—CORRELATIONS OF ARMY ALPHA AND GRADES RECEIVED IN THE FRESHMAN YEAR OF THE UNIVERSITY

	First term	Second term	Third term	Cases
Alpha and average grades	0.485	0.28	0.278	304
Alpha and English	0.517	0.42	0.29	265
Alpha and mathematics	0.213	0.47	0.02	94
Alpha and language	0.313	0.28	0.32	93
Alpha and history	0.540	0.07	0.09	37
Alpha and science	0.448	0.52	0.43	247
Average	0.419	0.34	0.234	

The next problem is that of discovering which one of the group tests is best for prophecying standings in school subjects. What is said is based on simple correlations. There might be combinations of tests within each group which would correlate more highly than any

[1] Jordan, A. M.: Some Results and Correlations of Army Alpha Tests. *School and Society*, March 20, 1920.

of the correlations mentioned below. That group test is defined as "best" from which may be obtained either from the group as a whole or from any individual test the highest correlations in the subject considered.

TABLE XII.—ARRANGEMENT OF CORRELATIONS TO SHOW HIGHEST CORRELATIONS OF EACH GROUP TEST WITH ALL SUBJECTS COMBINED AND WITH ENGLISH, MATHEMATICS, GENERAL SCIENCE, AND HISTORY

	All subjects	English	Mathematics	General science	History
Otis..........	5.......0.479 Group...0.450	4.......0.421 5.......0.425 6.......0.456 Group...0.466	5.......0.676 6.......0.526 Group...0.430	2.......0.513 5.......0.524 Group...0.502	Group....0.262
Army.........	5.......0.514 Group...0.476	3.......0.544 4.......0.527 Group...0.472	1.......0.608 2.......0.521 Group...0.511	2.......0.513 4.......0.514 5.......0.548 Group...0.456	3.......0.471 5.......0.421 Group...0.319
Miller.........	1.......0.466 Group...0.476	1.......0.594 Group...0.456	.... Group...0.456	1.......0.567 2.......0.531 Group...0.592	Group...0.168
Terman.......	1.......0.555 7.......0.492 Group...0.492	3.......0.572 7.......0.552 6.......0.492 8.......0.459 Group...0.436	5.......0.529 .... Group...0.466	1.......0.502 6.......0.516 7.......0.546 9 0.561 Group...0.636	6.......0.588 9.......0.429 Group... 0.408

In Table XII, "group" means all the elements of the test combined; the small numbers before the coefficients of correlation refer to the elements of the group test; thus, in the first line one would read "Otis, element five (arithmetical reasoning) has a correlation with all subjects of 0.479."

The inferences from Table XI are as follows:

1. For all subjects combined *Terman* stands above the rest because of Test 1 with a coefficient of 0.555, and because the correlation between the group of tests and all subjects is 0.492.

2. For English, *Miller* is best because by using Test I alone one gets a coefficient of correlation of 0.594.

3. For mathematics *Otis* is best since by using Test V alone a correlation of 0.676 may be obtained.

4. For general science, *Terman* is best since this group of tests gives an r of 0.636.

5. For history, *Terman* is best since Test VI gives a correlation of 0.588.

By using Terman's tests of intelligence, therefore, one gets the best test for general prognosis (0.55), best for general science (0.64), best for history (0.59), the second best for English (0.572) the third best for mathematics (0.529). The question of the value of these tests for general intelligence is not raised but only the question of prognosis for school work.

Discussion and Conclusion

Some tests are better for certain purposes than are others but in no case has a really high correlation been found. The highest correlation obtained is 0.68. I have calculated from a table of Thorndike's which is based on tenths that if we used fifths instead of tenths and wish to prophecy how many cases in each fifth of one series would be in the corresponding fifth in the related series, this number is 32 in 100 when the correlation is 0.5; the corresponding numbers when thirds are used are approximately 46 out of 100 with a correlation of 0.5. The figures for a correlation of 0.7 are, for fifths, 40 out of a 100; for thirds, 55 out of a 100; for a correlation of 0.9, 57 out of 100 for fifths, and 71 out of 100 for thirds. Thus it is seen that for reliable prognostic purposes we would need a correlation around 0.9. Which is manifestly impossible to obtain with grades since the self correlation of grades is not this high. These facts point to the distance we are from tests that will accurately prophecy. It may be that we shall be compelled to turn to the tests of specific abilities. The unexpected size of the correlations of some elements of these so-called general intelligence tests with grades also points as far as grades are concerned to the need of tests prognostic of standing in specific subjects such as the Rogers diagnostic tests for mathematics.

LANGUAGE ERROR TESTS[1]

(Concluded from the September issue)

G. M. WILSON

Boston University

The next problem was to bring these errors into new stories in satisfactory form and then re-test them to see whether or not there would result three compositions of approximately equal values. Again an advanced class was used in writing stories in an attempt to make use of these errors in satisfactory form for testing. With these as a basis, three tests were finally developed—Test A, "Saturday Morning;" Test B, "A Fishing Trip;" Test C, "An Accident." Through the cooperation of the superintendents of schools in Sioux City, Iowa, and Duluth, Minnesota, it was possible to try out these tests.

The combined medians for Sioux City and Duluth, based upon a total of 6965 tests, are shown in Table VII.

TABLE VII.—MEDIANS FOR SIOUX CITY AND DULUTH (COMBINED)

Grade	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Test A	10	17	20	22	23	23	25	25	25	26
Test B	15	13	19	20	21	22	25	26	27	26
Test C	17	15	20	23	22	24	25	26	26	26

The nature of the tests and their equality one with the other will be much more evident, however, from a showing of the total distributions. These are indicated herewith in Tables VIII, IX, and X. It is evident from these tables that Tests A, B, and C approach equality.

[1] The "Wilson Language Error Tests," will be published by the World Book Company, Yonkers, N. Y.

TABLE VIII.—DISTRIBUTION OF SCORES FOR RIGHTS
Test A. "Saturday Morning"
(Duluth and Sioux City combined)

Score	Grade										
	III	IV	V	VI	VII	VIII	IX	X	XI	XII	
0	1										
1	6	1									
2	4										
3	11	1									
4	10	1									
5	17	1									
6	21	3									
7	19	2	1								
8	15	4	3	1							
9	11	4									
10	17	6	2	...	1						
11	9	12	2	...	...	1					
12	14	13	5	1	2						
13	13	10	6	2	1						
14	11	13	5	5	1	1					
15	12	18	8	8	4	...	1				
16	11	15	12	3	4	1	...	...	1		
17	13	18	7	9	10	5					
18	8	11	14	14	13	6	2	1			
19	8	20	15	18	16	15	...	3	1		
20	5	15	20	20	24	20	2	4	1	1	
21	7	21	24	26	29	29	2	3	0	2	
22	1	6	31	36	23	37	4	12	6	2	
23	3	14	26	34	32	37	13	8	8	9	
24	...	6	16	49	43	47	8	16	8	8	
25	...	6	10	21	39	45	14	23	9	12	
26	2	1	8	8	23	33	15	31	13	18	
27	...	...	2	3	10	12	11	22	14	25	
28	...	...	3	...	4	5	...	4	5	6	
Total.....	248	222	220	258	279	294	72	127	66	83	1869
Q. 1........	6	13	18	20	20	21	23	24	23	25	
Median.....	10	17	20	22	23	23	25	25	25	26	
Q. 3........	15	20	23	24	25	25	26	26	27	27	

Table IX.—Distribution of Scores for Rights
Test B. "A Fishing Trip"

Score	Grade										
	III	IV	V	VI	VII	VIII	IX	X	XI	XII	
0	...	4									
1	...	4	1								
2	2	6									
3	1	3									
4	...	9	2								
5	1	8	2	1							
6	1	11	2	...	3						
7	6	5	1	2	3	...	1				
8	6	8	4	4	6						
9	13	5	6	2	5						
10	6	14	5	7	7						
11	14	12	8	3	3	1	2				
12	7	5	10	3	13	1	1				
13	20	10	5	8	12	1					
14	14	14	3	7	5	11					
15	16	5	12	10	19	7	1				
16	15	11	10	6	24	9	2				
17	16	9	17	8	29	14	1	...	1		
18	14	9	12	10	28	20	6	1			
19	9	10	12	8	31	24	5	...	..	1	
20	14	8	10	13	36	30	5	1	..	2	
21	6	9	15	8	39	40	11	2	3	3	
22	10	5	17	20	36	38	17	5	3	2	
23	5	6	10	23	36	29	19	10	3	7	
24	7	6	9	12	40	31	24	19	7	7	
25	4	3	14	8	33	30	36	16	8	16	
26	1	2	10	9	11	20	27	23	14	15	
27	...	5	4	8	22	14	27	26	23	10	
28	...	...	1	1	3	4	8	8	21	22	
29	...	...	2	...	3	4	8	4	6	4	
Total.....	208	206	203	181	447	328	201	115	89	89	2067
Q. 1........	12	8	15	15	17	19	21	24	25	25	
Median.....	15	13	19	20	20	22	25	26	27	26	
Q. 3........	19	18	22	22	24	24	28	27	28	28	

TABLE X.—DISTRIBUTION OF SCORES FOR RIGHTS
Test C. "An Accident"
(Duluth and Sioux City combined)

Score	Grade										
	III	IV	V	VI	VII	VIII	IX	X	XI	XII	
0											
1	2	...	1								
2	2	1									
3	1	4									
4	2	2	1	1							
5	4	5									
6	2	2									
7	1	6	3	1							
8	9	13	4	1							
9	4	6	4	1							
10	5	11	8	1	1	1					
11	6	16	10	...	1	2					
12	8	18	5	2	2	1					
13	8	10	4	5	2						
14	10	14	13	7	7	1	1				
15	10	14	7	4	2	3					
16	8	15	17	14	13	9	2				
17	10	17	15	11	15	7	2	2			
18	10	16	11	11	22	11	3	1			
19	10	9	19	17	20	19	5	1			
20	10	11	24	17	22	17	13	3	2	1	
21	16	10	10	17	27	26	10	5	4		
22	7	12	19	21	31	25	4	11	6		
23	7	9	24	34	49	35	19	16	8		
24	12	11	14	20	35	42	36	19	10		
25	9	4	23	27	34	36	31	16	17	2	
26	5	2	15	27	23	49	28	35	20	2	
27	1	1	15	15	12	29	36	27	26	2	
28	1	...	15	8	6	18	14	10	19		
29	...	...	5	9	1	9	2	10	7		
Total.....	180	239	286	271	325	340	206	156	119	7	2129
Q. 1.........	12	11	16	19	19	21	23	23	24	25	
Median.....	17	16	20	23	22	24	25	26	26	26	
Q. 3.........	21	20	25	25	24	26	27	27	27	27	

Further Evidences of Equality.—The limits of this article will not permit submitting extended data under this heading, but several things were done. All pointed strongly to the conclusion that the three tests are practically of equal value. The percentage of pupils in Grades III, VI, and X making each error in each test was figured. The median percentages for the different grades run surprisingly close together.

The 12-year-old group was separated and distributed. In all three tests, A, B and C, the median for the 12 year olds was 23 and the quartiles were either on the same point or only one step removed.

Tests B and C were given at the same time to pupils in grades III to VIII, inclusive, in a small city system. On the basis of 103 pairs thus secured, the coefficient of correlation was figured. The results showed a positive coefficient of 0.901.

On the basis of the above evidence, it seems safe to conclude that the three tests are, for all practical purposes, of equal value and may be used interchangeably. There are 28 errors in Test A and 29 in each of tests B and C, but the score for rights is practically the same for any of the three tests.

Probable Error of a Score.—Under the direction of Dr. Arthur S. Otis, the probable error of a score by the Difference Method[1] was figured. The resulting probable error was 2.24. This shows a high degree of reliability for the test. It confirms the high coefficient of correlation.

The Tests.—Tests A, B, and C in the form in which they were used follow herewith. It is planned later to revise and simplify the directions leaving the tests themselves in their present form:

Test A.—Correcting Language Errors. (*A Game*)

Name...Grade..............Age.........
Town......................School...........................Date................

Directions for the Game.—(To be read by the teacher, the pupils following.) This is a little *game* in which the pupil plays teacher, and corrects a composition written by a pupil. Correct by drawing a single line through words or expressions used incorrectly, and placing the correct words above them. For example, if you had the following sentence to correct: "He has went home" you would correct it by drawing a single line through *went* and writing *gone* above it. Make all changes necessary to secure correctness. Work at your usual rate. You will be given reasonable time in which to complete your work. When you have finished, turn

[1] Otis, Arthur S: Reliability of Binet Scale and Pedagogical Scales. *J. of Educational Research*, September, 1921.

the sheet right side down and leave it on your desk. All will be permitted to finish the work unless too slow.

The composition which you are to correct follows herewith:

Saturday Morning

Saturday morning is a busy time to are house. A feller has a good chance to work. Me and Dorothy we divide the tasks between us. Then we race to see who will finish first. Last Saturday I taken the breakfast dishes as one of my tasks. I am especial fond of washing dishes. You should have saw me work. I wanted to get through so as I could play. John he called up at eleven o'clock to see if I might play with him. I had two rooms to dust before I could go. John seen that I wouldn't leave my work until I had did all of it. He brought over some doughnuts and give them to me. I sure appreciated them doughnuts. Then John helped me. When we had finished, I suggested playing marbles until time for dinner. "I ain't got any marbles," said John. "They comes mighty handy," I replied. Then I give him some of mine. I had got to many for my bag. I and John enjoy playing marbles. When dinner was ready, mother invited John to stay. "If I was sure my mother wouldn't care, I would like to stay," he replied. John he seen that he was really wanted so he telephoned his mother. He enjoyed the dinner and ate heartily. When them apples were passed, John wanted one, but he couldn't eat no more. After dinner we had another game of marbles. I hope John may come over again.

Test B.—Correcting Language Errors. (*A Game*)

Name.......................................Grade............Age...........
Town...........................School..................Date..............

Directions for the Game.—(To be read by the teacher, the pupils following.) This is a little *game* in which the pupil plays teacher, and corrects a composition written by a pupil. Correct by drawing a single line through words or expressions used incorrectly, and placing the correct words above them. For example, if you had the following sentence to correct: "He has went home" you would correct it by drawing a single line through *went* and writing *gone* above it. Make all changes necessary to secure correctness. Work at your usual rate. You will be given reasonable time in which to complete your work. When you have finished, turn the sheet right side down and leave it on your desk. All will be permitted to finish the work unless too slow.

The composition which you are to correct follows herewith:

A Fishing Trip

John he is awful good to me. He once't ask me to go fishing and said that he could learn me to be a good fisherman on no time. He

had saw some men git a great many fish out of a deep hole about a mile up the river. He said that he had watched them until they got tired fishing. He seen them leave with a large sack full. I agreed to go with him. "We hain't got any bamboo poles," he said, "the folks haven't none left over from last year. Good poles is difficult to find." John give me the lunch to carry. We et our lunch before we done any fishing. I sit the table while John cut two poles and fastened the lines to them. He baited my hook hisself and told me to throw it in. I felt a bite at once and jerked the line. It was a large catfish. I was afraid the line would break. John said that the line was made good and had held to many big fish to break easily. I landed the fish but we didn't catch any more. We wanted to git another one so that each of us would have a fish to take home. We started and went home early. John said, "You can't never tell about beginners luck." John and I is good friends. As I neared home, I seen my little brother coming down the street. He had came to meet us. "I have got a big one," said I, as I showed him the fish. The fish was the main part of are supper that evening.

Test C.—Correcting Language Errors. (*A Game*)

Name..Grade..............Age........
Town...................................School...............Date..........

Directions for the Game.—(To be read by the teacher, the pupils following.) This is a little *game* in which the pupil plays teacher, and corrects a composition written by a pupil. Correct by drawing a single line through words or expressions used incorrectly, and placing the correct words above them. For example, if you had the following sentence to correct: "He has went home" you would correct it by drawing a single line through *went* and writing *gone* above it. Make all changes necessary to secure correctness. Work at your usual rate. You will be given reasonable time in which to complete your work. When you have finished, turn the sheet right side down and leave it on your desk. All will be permitted to finish the work unless too slow.

The composition which you are to correct follows herewith:

An Accident

One Friday afternoon are teacher she asked us if we wanted to go to the woods. It was an awful nice day! Ain't it fun to play in the woods on such days? Their was a woods near the school house. There was lots of flowers in bloom. John and me wanted to pick them flowers so's we could take some home to mother, but the others did not want to wait for us. We had not went far when we saw too squirrels. They run away from us. John ain't never seen such funny

little animals. He asked if he might throw a stone at them. He done it and the stone bounded back and struck him on the head. He had to pay up for it because their was a large bump on his head. We hadn't no medicine with us, so we had to go home to git some for him. If I were him, I would let the squirrels play next time. Me and William felt sorry for him. His mother give us some apples for bringing him home. There orchard was full of apples. They never had as many before. The apples was picked and lay in great piles under the trees. The apples which we received helped to make up for the disappointment in having to come home early.

It is hoped that superintendents and measurement experts throughout the country may make use of these language error tests, sending summary of the returns to the writer. Actually working with the tests and noting what can be done with them to determine the language ability of school children and the points on which a particular class is weak, will carry conviction that the tests are of real value.[1]

[1] The writer has been impressed as he has been using these language error tests that they have exceptionally high diagnostic value, and he hopes to set this forth in another article, showing at the same time just how to use and score the tests for diagnostic purposes.

THE RELATION OF RHYTHM TO THE HANDWRITING MOVEMENT

An Experimental Study

PAUL V. WEST

Assistant Prof. of Education, Univ. of Wisconsin

Much general interest has been attached in recent years to the problems relating to temporal controls of the handwriting movement. Chief among these problems are those having to do with rhythm. What is meant by the term "rhythmical writing?" Do better writers perform more rhythmically than poorer writers? What is the effect of imposed rhythm on performance? Is arm or finger movement better adapted to rhythmitization? What are some of the most pertinent suggestions governing the use of and the emphasis on rhythm in penmanship instruction?

Psychological experiments in rhythm have not been without significance and value in their general bearing on these inquiries, but direct experimentation in the field of handwriting is essential to satisfactory solutions. The investigations of Freeman[1] and Nutt[2] are outstanding reports which have contributed substantially to a better knowledge of rhythmic movement in handwriting. In brief they show that temporal rhythm in writing increases with age of the writer and is assisted by increase of speed. They also note that it is not correlated with good form, in fact has a tendency to interfere with quality in certain cases, and that there is little distinction, if any, in finger and arm movement as to influence on rhythmitization. These investigators have arbitrarily defined rhythmic movement as the tendency to give succeeding strokes the same time emphasis—in other words, rhythm is regarded as a simple uniformity in duration of strokes.

Two types of experiment here noted were undertaken with the desire primarily to gain more specific information regarding rhythmic movement as related to the penmanship of good and poor writers of both adult and child groups. The first of these was carried out by having the subject make to and fro movements, as in writing, with finger and arm movement, without and with imposed rhythm. By a system of electric contacts a kymographic record was made which

[1] Freeman, F. N.: The Handwriting Movement, *Educ. Mon.*, 1918, Vol. II, 2, 3.

[2] Nutt, H. W.: Rhythm in Handwriting, *El. Sc. Jour.*, Feb., 1917.

showed the relative amount of time spent on the stroke and the rest at the terminus of the stroke. The second experiment was effected by a photographic method by means of which the actual writing process was analyzed in terms of distance covered during each fiftieth of a second. This method proved very effective and accurate.

The first experiment did not provide means of studying actual writing movement as did the latter, but did reveal characteristic behavior of the respective sets of muscles used in handwriting. The criteria of rhythm here considered were (1) the approach to temporal uniformity of successive total strokes, from the beginning of each stroke to the beginning of its successor, and (2) equality of time spent on similar stroke elements (the period of movement or stroke proper, and the period of rest at the end of the movement) as well as their combination in the total stroke unit.

The results were analyzed so as to show the percentage of variation within these various units as summarized in Table I. The data of this table are presented graphically in Diagram I so as to make clear the effect of the imposed rhythm. The variation of the younger children is generally greater than that of any of the other groups.

Table 1.—Average Per Cent of Variability from Temporal Uniformity

Group		Rest		Stroke		Total		
		Finger	Arm	Finger	Arm	Finger	Arm	
Adult......	Good	13.7	18.2	10.5	8.7	6.8	6.4	Spontaneous
	Poor	15.4	13.6	11.0	7.9	6.3	6.5	
	Average	14.5	15.9	10.7	8.3	6.5	6.4	
Adult......	Good	14.3	16.8	10.8	9.7	8.0	7.1	Imposed
	Poor	16.1	17.3	13.0	11.0	8.4	7.8	
	Average	15.2	17.0	11.9	10.3	8.2	7.4	
Child......	Young	21.8	20.0	24.1	14.4	13.7	8.8	Spontaneous
	Older	16.3	15.8	13.0	9.4	14.4	7.9	
	Average	19.0	17.9	18.5	11.9	14.0	8.3	
Child......	Young	18.6	22.9	20.5	17.2	11.8	12.3	Imposed
	Older	15.8	21.7	15.5	14.9	8.8	9.7	
	Average	17.2	22.3	18.0	16.0	10.3	11.0	

Although there are many cases of overlapping the better adult writers maintain a slightly superior record of regularity. The rhythmic reaction to the total period as a unit is quite noticeable, for all groups here show less relative variation than in either the stroke or rest.

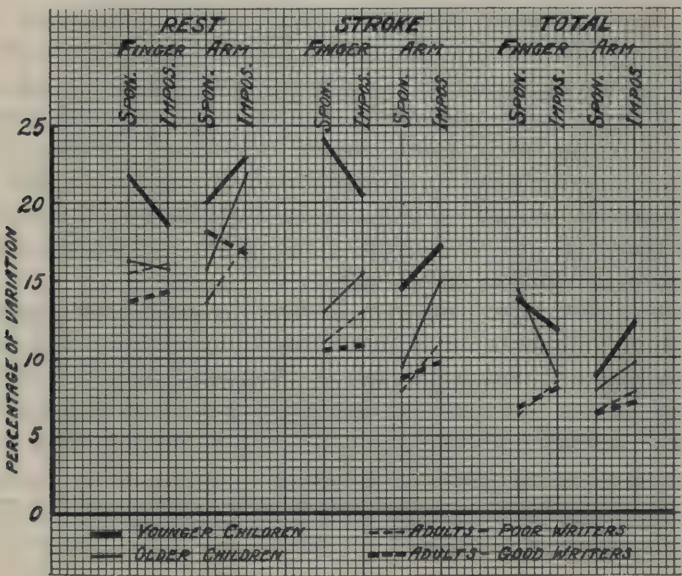


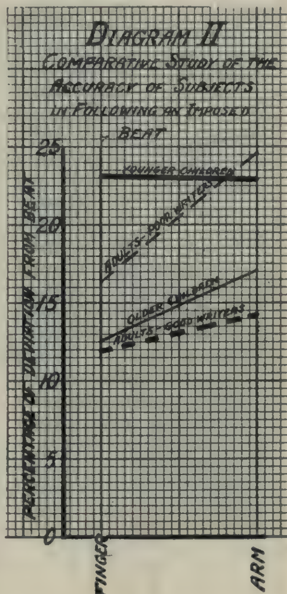
Diagram I.—Average Per Cent of Variability from Temporal Uniformity (Showing comparison of good and poor adult writers and younger and older children, under conditions of spontaneous and imposed rhythm, with the use of finger and arm movement.)

The finger and arm movements are doubtless distinctively significant for rhythmic behavior. The arm, because of its greater mass weight and consequent inertia, seems to be more easily adapted to natural rhythmic movement than the finger, especially in the case of children who have not yet gotten as ready control of the less gross finger mechanism. On the other hand, when the child attempts to follow a set rhythm, the inertia of the arm is manifest in the fact that the arm is less subject to rhythmitization than the fingers. Rhythmic habituation is apparently attained in arm movement, at least to marked extent, while the fingers, being unhabituated, yield more readily to the imposed beat.

This factor of habituation is marked also in comparison of the various groups as to the effect of an imposed rhythm on variation.

For adults, who have already acquired a comparatively high degree of rhythmic movement, probably as a more or less fixed norm of performance, the amount of variation is greater in both finger and arm movement when the rhythmic beat is followed than when the action is natural. The better writers show greater power of adaptation than the poor writers in this respect, however.

An investigation of the accuracy with which the imposed beat is followed reveals the fact that the poorer adult writers are less accurate than the older children, while the better writers show comparatively high ability to follow the rhythmic beat closely (Diagram II). In general the better writers and older children are able to anticipate the beat while the poorer writers and young children show a distinct tendency to lag behind the beat. The arm is found to be less accurate than the fingers in attaining the rhythmic pace set, except in the case of the younger children, who have not yet gained control of the finger muscles to the same degree as those who are older.



The imposed rhythm is found to increase the proportionate amount of time spent at rest for all groups, but this is especially notable in the case of the younger children, and when finger movement is used. The total duration of the stroke is increased, but the time spent on the stroke proper tends to remain constant.

In the second experiment the handwriting movement of 20 different subjects was investigated in a great deal of detail. Seven of the subjects were children ranging in ages from 7 to 16 years. Six of the remaining number, who were adults, were poor writers and seven were good writers as judged by the degree of evident coordination in their writing. Each subject wrote an indirect running oval and the majority wrote a sentence so that a study might be made of a simple repetitive form as well as a sample of characteristic penmanship.

In Figure I is given a reproduction of the written record of the 7-year-old child in making a running oval, with the successive 50th-second intervals indicated by the spaces between the cross bars.

The irregularity in form and in progression is evident. The poor organization of speed of movement is made very clear in Diagram III where the speed curve is in part compared with the curve of a record for the same exercise as made by a good adult writer. In the former there is no even progression of speed, and no adaptation of speed to the form to be produced, as in the latter. Also the many changes of speed are almost beyond comprehension and offer quite a contrast to the regular rise and fall of speed at the middle and end of strokes respectively which characterize the adult record.

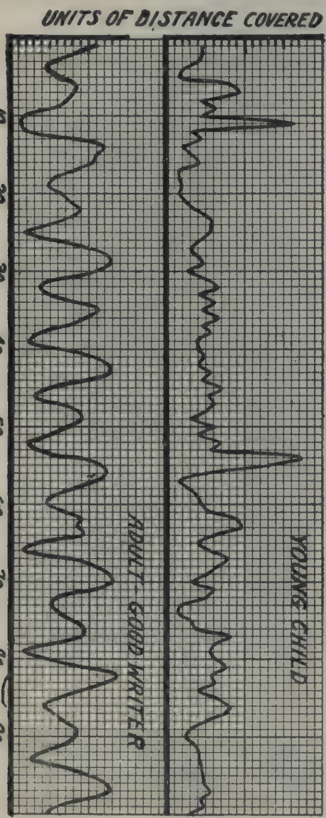


Diagram III.—Comparison of Speed Curves of the Young Child and Good Adult Writer in Writing the Running Oval.

Speed curves as thus constructed yield readily to rhythmic analysis since the number of units of time spent on each stroke can be accurately computed. An analysis of the records for the repetitive forms showed that the good writers among the adults made less variation from their average length of time both on the up stroke and down stroke as well as on the total of both than the poor writers, but very poor writers may show better natural rhythm than most of the good writers. The children, although in some cases showing a high degree of rhythm, are on the whole comparatively low in this respect.

When the records made in the writing of words were analyzed the rhythm was found to be very much lower, due undoubtedly to the irregularity in the length of strokes being constructed, as well as their complexity. While the individuals tend to give the same average length of time to strokes in the two writings of the same word or different words, the tendency is to devote unequal amounts of time to the respective strokes within a word. In general, the longer strokes receive the greater time emphasis, although this proportion is modified by the complexity of the stroke.

Ovals were analyzed to see what was the effect of curvature on the speed attained within the stroke. In Figure II is shown a portion of an oval broken up for measurement into seven arced segments having their centers of radiation in order as indicated by the numerals 1 to 7. It is notable that the radial lengths differ greatly. In the complete form of this particular oval the correlation between radial length and

FIGURE ONE
Writing of Seven Year Old Child Showing
Comparative Distances Covered in each 50th of a Second.

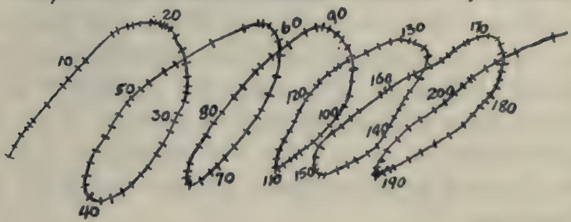
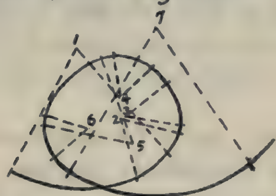


FIGURE TWO
Showing Method of Analyzing Curve
into Arced Segments.



maximum speed within the arc for 33 segments was found to be very high by the Spearman Rank Method, the coefficient being 0.96. The line of regression however indicates a curved rather than a straight line relationship with a falling off in the regular increment of speed after a certain radial length has been reached, and a tendency for the rate to reach a fixed limit as the arc approaches a straight line form. This relation was found to be quite generally characteristic of the subjects. The poorer writers and children show very much lower correlation between the maximum speed and curvature of the arc than do the good writers and also exhibit a more marked tendency to slow down the speed relative to the increase of the radial length of the arc.

If, as the strokes decrease in curvature, the maximum speed within the stroke were to remain constant, the time must be the variable, varying directly with the radial length of the stroke in case the stroke

resembles a simple arc. But were the speed to vary in exact proportion as the radial length the time would remain constant. The closer approach to this constancy on the part of the good writers indicates a marked tendency toward temporal rhythm. If the strokes used in writing could be so made as to be perfectly arced segments, it is possible that they might be constructed in about the same length of time regardless of size or degree of curvature. But the form of the written word does not ordinarily yield to such simple temporal analysis.

Experimentation with a few subjects in the following of a rhythmic beat while writing words showed that a writer with habituated speed and rhythm would be slowed down by any beat that was slow enough to be consciously followed. The regularity of time thus induced in the successive strokes was also in marked contrast to the normal behavior of the subject, or of any one else examined, and resulted in uncertainties of speed progress which affected harmfully the quality of the written form.

The fluency, freedom and smoothness of movement which characterizes the good writer generally is the fundamental "rhythm" of handwriting. It is due to the unhampered interaction and coordination of the fingers and arm as free swinging levers, and with a ready control of the motor elements at every instant, so there is a definite relation between the form being constructed and the speed and time relationships.

The use of the rhythmic guide in penmanship instruction must be carefully supervised and scientificaly directed. The value of such a guide in the early years is unquestioned, as an aid in organizing the writing into unit strokes, and also as an aid in facile movement. For purposes of temporal rhythmitization of movement an imposed rhythm is not thoroughly adaptable to the process of constructing word forms, but may be used to better advantage with simple repetitive forms. After the subject has developed habits of natural rhythm the use of the rhythmic guide may result in a retardation of speed and a detrimental effect on form. Little is known as to the extent to which an imposed rhythm may be acquired as a fixed habit by practice. It appears that its chief value must lie in the development of a smoothly regulated type of movement in which the writer has a controlled adjustment of speed to form. The training process must involve due consideration of all the muscles naturally used in handwriting, as well as all individual differences due to age and natural tendencies to rhythmic movement.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

New Studies in Reading

A Consideration of Essential Principles of Teaching Reading and Literature.—Professor Leonard's new book[1] on reading is an exceedingly stimulating and constructive volume embodying a point of view which issues from a sound conception of educational values. Mr. Leonard has drawn heavily on his experiences to illustrate and clarify his discussion and where that did not suffice he quotes generously from the experience of numerous co-workers. The abundance and quality of these "samples" bridge the gap so often left between theory and practice and makes the book as readable as it is instructive. The discussion of tests and measurements are illustrated by numerous well chosen graphs. Illustrations depict actual school dramatizations and library conditions. One appendix contains a bibliography related to each chapter of the book; another sixty-three page appendix of annotated booklists for children is arranged by grades as well as by subjects and covers all grades. The book is very thoroughly indexed.

L. Z.

---

3. *A Critical Study of the Content of the School Reading Course.*—Critics of the content of the elementary school curriculum in reading will appreciate the outcomes of Dr. Uhl's study[2] based on the reactions of over 2000 teachers in more than 100 cities and on an analysis of over 500 pupil scores and judgments. These were used in constructing standards for rating reading selections[3] proposed for use in the selection, elimination and grade placement of materials.

The method of the investigation and the treatment of data lead

---

[1] Leonard, Sterling A.: "Essential Principles of Teaching Reading and Literature." J. B. Lippincott Co., Philadelphia, 1922, p. 460.

[2] Uhl, Willis L.: Scientific Determination of the Content of the Elementary School Course in Reading. "University of Wisconsin Studies in the Social Sciences and History," No. 4., Madison, 1921, p. 152.

[3] Rating Scales for Reading Selections.

to significant findings which are clearly set forth in tabular and graphical form. The representation of changes in pupils' interest in and comprehension of certain selections provides a new approach to grade placement and should be reflected in a happier and wiser choice of literary matter for use in the grades.

Interest and difficulty are considered separately in the rating scales with which the study concludes, and each value is illustrated by several well known samples.

L. Z.

---

4. *Reading Research Interpreted for Teachers.*—Dr. C. T. Gray has sought to restate the results of his own investigations and others in a volume designed for use in special methods courses and reading circles.[1] In no case does he dogmatize without presenting scientific evidence. Reading ability is analyzed from four standpoints and the possibilities of diagnosis are pointed out from the standpoint of standard tests, perception, motor elements and the higher mental activities of reading. This analysis and synthetic summary comprise the 14 chapters of Part I.

Part II discusses the methods and materials necessary for diagnostic observation and measurement by means of which the presence and degree of specific deficiencies may be noted. Five cases are described in connection with interpretation of a Diagnostic sheet. The two brief chapters in Part III are devoted to a discussion of the principles relating to remedial measures, followed by a bibliography of 20 numbers and an appendix on statistical methods.

L. Z.

---

5. *Growth Curves Obtained from an Objective Study of Reading Habits.*—Some of the limitations of standardized tests are aptly demonstrated by the type of measurement undertaken by Dr. Buswell[2] and reported in the second Chicago monograph over his signature.

---

[1] Gray, C. T.: "Deficiencies in Reading Ability." D. C. Heath and Co., Boston, 1922, pp. XIV + 419.

[2] Buswell, Guy T.: Fundamental Reading Habits: A Study of Their Development. *Supplementary Educational Monographs*, No. 21, University of Chicago, Chicago, 1922, pp. XIV + 150. $1.50.

The development of fundamental eye habits is traced through the various growth stages from earliest attempts to read in the first grade to the maturity exemplified by college seniors. In a detailed analysis of first grade reading records made at stated intervals by the use of a further improvement on the apparatus used by C. T. Gray and others, various approaches to reading are contrasted and compared psychologically.

In order to keep other factors constant all subjects above the first grade read the same material. For each of 179 cases the records show the average number of fixations per line, average duration of eye-pauses and average number of regressive movements per line. Growth curves show the development in each of the elements listed above. Individual variations resulting from various cases are an index of needs which specific training should supply. The author points out the danger of evaluating any method from a skill measure taken at an early growth stage and shows that devious paths of growth may lead to maturity, although some necessitate needless meanderings. This study should lead to further research and did in the scientific selection of procedure. The attempt to get objective indices of growth in *habits* and attitudes is encouraging to those who count such values at least as significant as skill.

L. Z.

---

6. *A Case Book in the Reading Field.*—Clinical Study as a method of education and pursued by several members of the Chicago group. Paralleling the technic of medical research and social investigation Dr. W. S. Gray[1] and his co-workers studied 27 individual cases to determine the significant characteristics of poor readers, causes of difficulty, and appropriate remedial instruction. A study of each child's history was made by compiling available school records, and consulting teachers and parents. In addition to an analysis of the results of standardized reading tests and the use of intelligence tests the specific nature of each child's difficulty was painstakingly determined by the aid of special tests, the use of laboratory equipment, analysis, observa-

[1] Gray, William S. with the cooperation of Delia Kibbe, Laura Lucas, Lawrence W. Miller: Remedial Cases in Reading: Their Diagnosis and Treatment. *Supplementary Educational Monographs*, No. 22, University of Chicago, Chicago, 1922, p. 208. $1.75.

tion and interview. Individual remedial instruction was organized under carefully controlled conditions. After approximately 2 months of prescribed training during which individual reactions were recorded, the pupils were re-tested.

The monograph reports before and after scores of each case together with a detailed record of procedures. Cases are grouped with reference to the type and cause of difficulty. Pupils who make little or no progress in reading are most easily detected but the causes for such deficiency were found hardest to determine because of the variety of possible contributing factors. Such pupils need a special type of instruction. It is highly desirable that they be recognized early, because their disability is often due to physical or intellectual limitations.

In the next group are included all pupils who encountered serious difficulty in interpretation. Although the causes for such deficiency were found to be numerous and the remedial procedure had to be varied to suit the conditions of each case, it is significant to note that effective instruction resulted in the removal of the cause.

Pupils who encountered difficulty in the mechanics of reading, form the third group. Ten separate causes were isolated and effective remedial measures were necessarily different and related to particular causes. Pupils whose rate of silent reading was unsatisfactory are next considered. Gray concludes that undue emphasis on rate before rudimentary habits are established may easily increase difficulties of recognition and prevent some pupils from becoming fluent readers.

Remaining cases are put together as weak in all phases of reading. In such cases individual remedial work is especially urgent and helpful if based on a recognition of the varied needs.

The monograph concludes with a brief chapter on similar studies carried on in a city school system and will no doubt be instrumental in bringing the possibility and value of such work to the attention of other progressive communities.

Educational practitioners will find Dr. Gray's case book suggestive in the diagnosis of thousands of other cases in which the "symptoms" are somewhat similar. Cumulative records of results add, determine, and reveal the validity of remedial measures and provide new data.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Vol. XIII November, 1922 No. 8

THE LIMITS SET TO EDUCATIONAL ACHIEVEMENT BY LIMITED INTELLIGENCE[1]

MARGARET V. COBB

Institute of Educational Research

Teachers College, Columbia University

I. Previous Estimates of the Limits of School Progress

Very few figures are available which show definitely the maximum school progress that is possible for children of any given level of intelligence, either in terms of rate of learning, or in terms of the upper academic limit of their achievement. Dr. Terman (1), in connection with the description of the Stanford-Binet Scale, gives general statements the substance of which is that children below 75 IQ should be kept out of regular classes, and will rarely be equal to the work of the fifth grade, however long they attend school; that for children below 80 IQ, special classes are advisable and work should be concrete; that children 80 to 89 IQ will usually be able to reach Grade VIII after from one to four failures; that those from 90 to 109 may enter high school, but marks will be below average there and excessively poor in college; that those from 110 to 119 should complete eight grades in 7 years and are "good" scholars in the grades, average in high school. Children who stand better than 120 IQ, he says, are so intelligent as to be seriously hampered in an ordinary classroom. Later, in "The Intelligence of School Children," (2) he says, "Throughout Proctor's study it appears that the standards of work which are maintained in the first year of average California high schools can not be satisfactorily met by pupils with a Stanford-Binet mental age below 13 years, and that below the mental age of 14 years the chances of success are not

[1] The investigation reported here was made possible by a grant from the Commonwealth Fund, and was carried out under the guidance of Dr. Thorndike.

good . . . Entrance to this high school is pretty well barred to children who test much below 90 . . . In this high school, at least, the pupil with IQ below 90 is practically certain to fail in such studies as algebra and Latin . . . Below 90 IQ graduation is by no means likely."

Proctor (3), in connection with his study of the usefulness of intelligence tests in giving guidance to high school pupils, found on examination of the high school success of 107 freshman, that 70 per cent of those who tested below 95 in intelligence quotient on the Stanford-Binet Scale (19 pupils) failed in more than half of their subjects. All of this 70 per cent (13 pupils) either dropped out of school or repeated.

The annual report of the Providence Public Schools for the year 1917–1918 (4) contains similar information. In summary of the work in mental measurements of children, it gives in tables and in general statements the following information concerning the grammar schools:

The most usual mark of children whose intelligence quotients (Stanford-Binet Scale) are between 70 and 79 is *D* (conditioned), with more children failing (*E*) than passing (*C*). Only one child, among the 220 included in the table whose intelligence quotients were below 80, made a better mark than *C*. Among those whose intelligence quotients range from 80 to 89, the most usual mark is *C*, with more conditioned or failing than there are who do better than *C*. Among the group whose intelligence quotients range from 90 to 99, the most

Table I.—Correlation of Intelligence and School Work

IQ	Failed, *E*	Conditioned, *D*	Passing, *C*	Good, *B*	Excellent, *A*	Total	Distribution of intelligence
Below 70	27	36	4	...	..	67	0.055
70–80	36	92	24	1	..	153	0.15
80–90	8	71	125	21	2	227	0.223
90–100	1	22	133	40	3	199	0.195
100–110	1	10	67	77	15	170	0.167
110–120	..	4	23	66	23	116	0.114
120 and above	..	...	12	32	40	84	0.082
Total distribution, number..........	73	235	388	237	83	1016	
Marks, per cent.....	0 071	0.231	0 381	0.232	0.081		

frequent mark is still *C* but with rather more exceeding this than there are below it. Among the group from 100 to 109 there are more children making *A* and *B* than there are making grades below *B*. Above 109, no children fail and only 3½ per cent are conditioned, while most of them make *A* and *B*. (In connection with Table I it should be explained that it does not represent the Providence schools, since more of the backward than of the normal children were included.)

A second table in this same report includes an estimate as to the probable limit of successful school progress of children having different intelligence quotients. For comparison, Terman's estimates, from "The Measurement of Intelligence" (1), have been added to the table. (See Table II.)

Table II

IV	Mental age at 14	Probable limit of school progress	
		Providence estimate	Terman estimate
60–70	9-8	V*A*	IV*A*
70–80	11-2	VII*B*	VI*A*
80–90	12-6	VIII*B*	VIII*B*
90–100	14	VIIIA	
100–110	15-4	VIIIA	High school
110–120	16-8	VIIIA	High school (average success)
120–150	18-2 (130) 19-6 (140) 21 (150)	First year high or more	

When the Providence figures are compared with the estimate given by Terman in "The Measurement of Intelligence" (1), it will be seen that the discrepancies are not very great, the estimates being identical at some points and never more than a year apart. The Providence table would seem to indicate that in order to do successful work in the first year of high school, an intelligence quotient of 120 is necessary. Terman would predict average high school success where the intelligence quotient is 110.

A third estimate is that of Supt. Carroll, who considers that an intelligence quotient of 110–115 is necessary for high school work.

These data obtainable this year have been chiefly in terms of the

Army Alpha Examination. In the following sections our figures will be expressed, so far as possible, both as Alpha scores, and in terms of the Stanford-Binet Scale. The following table, made up after a study of the Army (5), Kohs-Proctor (3), Kansas State Normal School (6), and Doll (7) norms, is the one which we have used, though naturally it is far from being considered final.

TABLE III

Median MA	Median Alpha score	Median MA	Median Alpha score
20-6	175	14-6	85
19-6	160	13-6	70
18-6	145	12-6	55
17-6	130	11-6	40
16-6	115	10-6	25
15-6	100	9-6	10

II. The Intelligence of High School Pupils

The Army Alpha examination has been rather extensively used in high schools. In Kansas, the Bureau of Educational Measurements and Standards of the Kansas State Normal School at Emporia has collected data from a number of high schools (6). The medians from this survey are included in our tables and the distributions themselves (given by them in terms of letter grades) are similar to those here presented. Madsen and Sylvester (8) presented in "School and Society" in 1919 distributions obtained in three high schools in Illinois, Iowa and Wisconsin. The Bureau of Educational Reference and Research of the University of Michigan (9) issued in 1921 medians for some schools in that state; these schools[1] have cooperated by supplying their records for this study, and the distributions (see Tables IV–VII and Figure 1) are here given separately as well as in combination with Madsen's published figures.

Michigan may apparently be taken as fairly representative of the intelligence of the country as a whole. Reference to Section V will show that recruits from Michigan, and medical officers from Michigan, m de scores on Examination Alpha which were very slightly above

[1] Alma, Milan, Mt. Clemens, Mt. Pleasant and Detroit.

those of the country as a whole. While therefore it should be borne in mind that the group here reported is a local group, one will not go far amiss in generalizing its results.

Most striking is the wide variety of intelligence to be found in any one high school year. A freshman may score anything from 35, corresponding approximately to a mental age of 11 years, to 185, which is a very superior adult score. Few scores however (7 or 8 per cent) fall below 65 (MA about 13 years), and some of these are probably scores which are low through some accident. The median score of recruits to the National Army was 63 (MA about 13 years).

Whether the army was typical in intelligence of the total population of the United States is a question still under discussion. At both ends of the scale are groups in the general population which did not get into the army—at one end the feeble-minded, at the other, intelligent men exempted because of the importance of the work they were carrying on in civil life. Terman (10) believes that the recruits were on the whole lower than the whole population; Doll (7) and Goddard (11) consider the army a representative group. At any rate it is the best sample we have or are likely to have for some time of the intelligence of this country, and with the above cautions in mind, army figures will henceforward be used for reference as standing for the population as a whole. For this purpose officers in proper proportion[1] (see Figure 1) have been included with the recruits; the median for this total distribution is 65.

These results may well be compared with the Army figures (5), which indicate, in general, conditions which existed 10 years or more ago, both as to the proportion who continued in school and the stiffness of the requirements they had to meet. With this allowance, the Army figures are not widely discrepant from the results of Army Alpha given in schools of the present day. The difference which appears between the figures for recruits and those for officers shows up in a very interesting way the importance of the part which may be played by other personal qualities than intelligence in determining continuance in school and a commanding position in the world.

From these Army figures it appears that it was even more rare 10 years ago than it is today for a man with less than average Alpha score

[1] The proportion used was approximately one officer to 15 recruits. This is so small a proportion that the effect on the total distribution, while appreciable, is not great; it would have made no significant difference had the proportion used been 1 to 12, or 1 to 18 or 20.

TABLE IV.—ALPHA DISTRIBUTION OF HIGH SCHOOL FRESHMEN

Alpha score	Mt. Clemens	Milan	Mt. Pleasant	Alma	Mich. outside Detroit	Ill., Wis., and Ia.	Total, Ill., Wis., Ia., and Mich.	Per cent
205								
200								
195								
190								
185								
180	.....	.....	....	......	......	3	3	0.17
175	.....	.....	....	.....	.....	0	0	
170	.....	.....	....	.....	.....	2	2	0.12
165	1	.....	....	.....	1	2	3	0.17
160	.....	.....	1	1	2	5	7	0.41
155	.....	.....	....	.....	.....	7	7	0.41
150	.....	1	....	2	3	9	12	0.70
145	1	2	....	......	3	13	16	0.93
140	1	.....	....	2	3	17	20	1.16
135	1	2	2	4	9	31	40	2.32
130	1	2	2	3	8	50	58	3.37
125	4	1	2	6	13	67	80	4.65
120	3	3	2	7	15	65	80	4.65
115	6	2	3	11	22	72	94	5.46
110	6	4	2	10	22	73	95	5.52
105	8	2	8	5	23	97	120	6.97
100	5	6	8	12	26	87	113	6.57
95	7	6	5	27	45	112	157	9.13
90	4	7	10	11	32	98	130	7.55
85	8	11	6	18	43	113	156	9.07
80	3	7	11	10	31	88	119	6.92
75	5	7	9	18	39	73	109	6.34
70	6	14	8	11	39	61	99	5.75
65	5	8	6	11	30	37	71	4.13
60	4	4	7	12	27	23	50	2.90
55	3	1	3	6	13	16	29	1.68
50	1	4	....	8	13	15	28	1.63
45	2	3	1	3	9	5	14	0.81
40	.....	1	....	.....	1	0	1	0.06
35	.....	1	....	1	2	4	6	0.35
30	.....	.....	....	.....	.....	0	0	
25	.....	.....	....	.....	.....	2	2	0.12
20								
15								
10								
No. cases	85	99	91	199	474	1247	1721	
Median.....	96.07	84.64	85.4	90.68	88.84	98.95	96.48	
Median... MA.....	15-3	14-6	14-6	14-10	14-9	15-5	15-3	

TABLE V.—ALPHA DISTRIBUTION OF HIGH SCHOOL SOPHOMORES

Alpha score	Mt. Clemens	Milan	Mt. Pleasant	Alma	Mich. outside Detroit	Ill., Wis., and Ia.	Total, Ill., Wis., Ia , and Mich.	Per cent
205								
200								
195								
190								
185								
180	......	....	.....	.....	......	2	2	0.16
175	......	....	.....	.....	......	1	1	0.08
170	......	....	.....	.....	......	6	6	0.48
165	2	....	.....	.....	2	7	9	0.79
160	1	....	1	.....	2	7	9	0.79
155	......	....	.....	2	2	15	17	1.36
150	......	1	.....	1	2	20	22	1.76
145	4	....	1	4	9	30	39	3.11
140	4	....	.....	.....	4	47	51	4.07
135	......	....	4	.....	4	34	38	3.03
130	6	2	4	.....	12	63	75	5.90
125	6	5	2	2	15	67	82	6.54
120	3	0	8	4	15	76	91	7.26
115	7	3	2	9	21	72	93	7.42
110	5	1	5	5	16	94	110	8.78
105	7	6	3	7	23	70	93	7.42
100	3	3	5	6	17	83	100	7.98
95	2	5	4	9	20	65	85	6.78
90	5	5	9	9	28	67	95	7.58
85	1	5	5	7	18	50	68	5.42
80	4	3	3	7	17	30	47	3.75
75	1	2	5	5	13	31	44	3.51
70	......	1	2	3	6	22	28	2.23
65	......	....	3	2	5	16	21	1.68
60	......	....	2	3	5	8	13	1.04
55	......	2	.....	2	4	1	5	0.40
50	......	1	1	.....	2	3	5	0.40
45	......	....	.....	.....	......	2	2	0.16
40	.....	....	.....	.....	......	0	0	......
35	......	....	.....	.....	......	1	1	0.08
30						0	0	......
25						0	0	......
20						0	0	......
15	......	....	.....	.....	......	1	1	0.08
No. cases	61	45	69	87	262	991	1253	
Median....	116.79	98.5	100.5	98.06	103.82	112.42	110.84	
Median ... MA.....	16-7	15-5	15-6	15-5	15-8	16-3	16-2	

Table VI.—Alpha Distribution of High School Juniors

Alpha score	Mt. Clemens	Milan	Mt. Pleasant	Alma	Mich. outside Detroit	Ill., Wis., and Ia.	Total, Ill., Wis., Ia., and Mich.	Per cent
205								
200								
195								
190								
185	...	..	...	...	......	2	2	0.20
180	...	..	1	...	1	3	4	0.41
175	...	..	...	2	2	5	7	0.72
170	2	..	1	...	3	8	11	1.13
165	1	..	2	...	3	15	18	1.84
160	1	..	...	...	1	15	16	1.64
155	4	..	2	1	7	32	39	3.99
150	2	..	2	...	4	38	42	4.30
145	2	..	2	3	7	34	41	4.20
140	2	..	4	4	10	51	61	6.24
135	1	..	2	5	8	60	68	6.96
130	3	..	...	3	6	62	68	6.96
125	4	1	4	1	10	66	76	7.78
120	5	..	3	2	10	74	84	8.60
115	5	..	7	4	16	76	92	9.41
110	3	..	2	6	11	68	79	8.09
105	3	..	3	2	8	52	60	6.14
100	3	1	3	3	10	48	58	5.94
95	1	1	2	7	11	42	53	5.42
90	1	..	3	2	6	23	29	2.97
85	1	..	1	4	6	19	25	2.56
80	...	..	1	3	4	13	17	1.74
75	1	..	1	2	4	7	11	1.13
70	...	..	...	3	3	3	6	0.61
65	...	..	...	1	1	3	4	0.41
60	...	..	...	2	2	0	2	0.20
55	1	..	...	2	3	0	3	0.31
50	...	..	...	...	......	0	0	
45	...	..	...	...	......	1	1	0.10
40								
35								
30								
25								
20								
No. cases	46	3	46	62	157	820	977	
Median...	124		120	110	117.97	123.72	122.89	
Median MA....	17-1		16-10	16-2	16-8	17-1	17-0	

TABLE VII.—ALPHA DISTRIBUTION OF HIGH SCHOOL SENIORS

Alpha score	Mt. Clemens	Milan	Mt. Pleasant	Alma	Mich. outside Detroit	Ill., Wis., Ia.	Total, Ill., Wis., Ia., and Mich., outside Detroit	Per cent	Detroit[1] Number	Detroit[1] Per cent
205										
200	......	..	.....	.....	......	1	1	0.13		
195	......	..	.....	.....	......	0	0			
190	......	..	.....	.....	......	2	2	0.26		
185	......	..	.....	.....	......	0	0			
180	1	..	.....	1	2	2	4	0.52	3	0.48
175	1	..	.....	.....	1	5	6	0.78	3	0.48
170	1	..	.....	1	2	12	14	1.83	11	1.77
165	......	..	1	2	3	15	18	2.35	6	0.97
160	......	..	2	1	3	18	21	2.74	17	2.74
155	1	..	3	3	7	31	38	4.96	19	3.06
150	2	..	2	3	7	26	33	4.30	30	4.83
145	......	..	3	2	5	36	41	5.35	41	6.60
140	2	..	4	3	9	51	60	7.83	34	5.47
135	1	..	4	1	6	41	47	6.13	41	6.60
130	1	..	3	1	5	45	50	6.52	41	6.60
125	1	..	4	3	8	59	67	8.75	50	8.05
120	4	1	4	7	16	44	60	7.83	44	7.08
115	2	..	2	2	6	55	61	7.96	48	7.72
110	2	..	5	4	11	49	60	7.83	46	7.40
105	2	..	1	6	9	42	51	6.66	49	7.89
100	1	..	1	6	8	34	42	5.48	32	5.15
95	1	..	2	1	4	18	22	2.87	33	5.31
90	2	..	2	1	5	15	20	2.61	19	3.06
85	1	..	3	1	5	4	9	1.17	22	3.54
80	......	..	.....	1	1	7	8	1.04	8	1.29
75	......	..	2	3	5	15	20	2.61	11	1.77
70	1	..	.....	......	1	3	4	0.52	4	0.64
65	......	..	.....	1	1	3	4	0.52	3	0.48
60	......	..	.....	......	0	1	1	0.13	2	0.32
55	1	..	.....	......	1	1	2	0.26	3	0.48
50	......	..	.....	......	......	......	......	....	1	0.16
45										
40										
35										
30										
No. cases	28	1	48	54	131	635	766	....	621	
Median	121.25	..	127.5	120.71	122.66	127.25	126.42	....	123.35	
Median MA...	16-10	..	17-4	16-10	17-0	17-4	17-4	....	17-0	

[1] The median scores of the seniors in seven different high schools in Detroit were as follows:

SCHOOL	MEDIAN	MA	NO. CASES
Northeastern..................	115.0	16-6	108
Eastern......................	118.75	16-9	81
Southeastern..................	120.0	16-10	92
Central......................	123.33	17-0	114
Western......................	125.0	17-2	56
Northwestern..................	128.96	17-5	87
Northern.....................	128.21	17-5	83

to have entered high school. High school pupils of that day, like present-day high school pupils, had at least average intelligence, as thus measured; in other words, the lower half of the population (in Alpha intelligence), practically without exception, had not continued in school beyond the elementary grades.

It will be seen then that in spite of the wide range of ability in high school freshmen, of which every high school teacher must be conscious, there are nevertheless few (less than 7½ per cent) who have not median intelligence or better, as measured by the Alpha examination. This corresponds approximately to a mental age of 13. The pupils in academic high schools are, in fact, a limited group, which just about covers the upper half of the whole range of American intelligence.

This is significant. Each year a larger and larger proportion of school children has been going on into high-school, until at present about half of the children who enter the first grade may be expected eventually to enter high school. (Local variations in this proportion are great. See Section V.) Since many of the more intelligent children still do not have the chance to go to high school, it will very soon be true that an appreciable per cent who are below average will be attempting the course. Thus the question of the degree of intelligence which is essential for success in the older academic course, and in the newer vocational courses (commercial, manual training, household economics, etc.) becomes of first-class importance.

III. Continuance in School in Relation to Intelligence

One way to get a very general look at this problem is to consider continuance in high school in relation to intelligence. It is obvious from the figures that sophomores do better than freshmen in the tests, juniors than sophomores, and seniors still better. Is this because the less intelligent pupils have found the way too hard, and have dropped out? Or is it be explained by the mental growth of the children, and the additional information and skill which they have acquired? We may grant at once that both causes are at work; improvement in scores (above what practice brings) appears when the same children are tested as freshmen, and again as seniors; while of a given freshmen class those who drop out before senior year are somewhat more largely from the lower than from the upper half of the distribution. (See Cobb and Tape in "School and Society"(12).) It is not yet

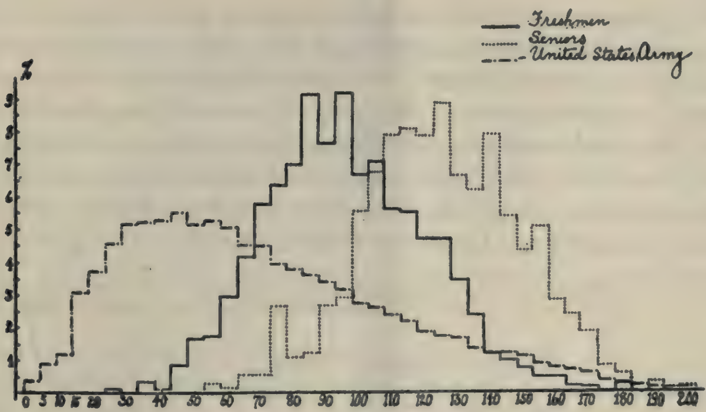


FIG. 1.—Distribution of Alpha scores of high school freshmen and high school seniors compared with Alpha scores of literate recruits and officers of the United States army.

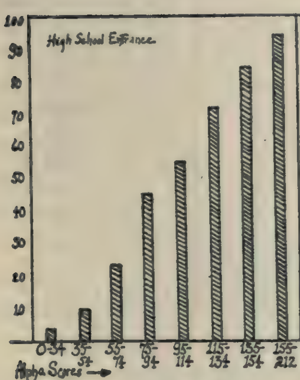


FIG. 2.—Per cent of recruits who had entered high school at different Alpha levels.

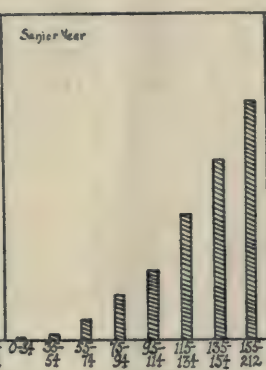


FIG. 3.—Per cent of recruits who had become high school seniors at different Alpha levels.

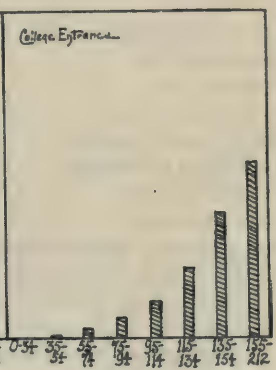


FIG. 4.—Per cent of recruits who had entered college at different Alpha levels.

possible to decide just how much of the yearly gain to ascribe to each of these factors. Growth, instruction and added experience are perhaps to be credited with more of it than is elimination of the less intelligent; but all we can say quite surely is that each plays an important part. This yearly gain, which is about 15 points at first, and decreases somewhat, is still found when high school seniors are compared with college freshmen. In Table VIII several college groups have been added for comparison; the gain is about 5 points.

Table VIII.—Median Alpha Scores, High School and College Groups

	Freshmen	Sophomores	Juniors	Seniors	College freshmen
Mt. Clemens	96.07	116.79	124.0	121.25	
Milan	84.64	98.5			
Mt. Pleasant	85.4	100.5	120.0	127.5	
Alma	90.68	98.06	110.0	120.71	
Michigan (outside Detroit)	89.05	103.82	117.97	122.6	
Illinois, Iowa and Missouri	98.95	112.42	123.72	127.25	
Illinois, Iowa, Missouri and Michigan	96.65	110.84	122.89	126.42	
Detroit	......	......	......	123.35	
New York (Kansas report)	92.0	104.0	118.0	132.0	
Emporia, Kan	80.0	105.0	101.0	111.0	
Stanton, Va	91.0	114.5	136.0	117.0	
Kansas report	......	......	......	......	129.0
Ohio State University	......	......	......	......	130.0
University of Illinois	......	......	......	......	131.0
Oberlin College	......	......	......	......	148.4
Yale University	......	......	......	......	159.7
Recruits who had entered high school (tested 15 years out of school on the average)	97.76	104.61	111.36	115.06	118.7
Officers who had entered high school (tested 15 years out of school on the average)	140.68	141.34	141.97	142.55	143.3

This relation of intelligence to continuance in school may be brought out also in another way. The Army figures show very definitely that at the time these recruits were of high school and college age, say 5 to 10 years ago, the more intelligent youths all along the line

remained longer in school than those who made lower scores. Table IX shows the per cents entering high school, college, etc., at each Alpha level. Figure 2 illustrates the first column of this table, and shows the situation for high school freshmen. Of those scoring less than 35 on Alpha, 4 in 100 reported that they had entered high school; of those scoring 155 or better, 92 in 100, or 23 times as many, so reported. This comparison, based on Alpha scores, omits the illiterate group altogether. It is probable that, had they been included, the chance of entering high school would be at least 30 times as great for those over 155 as for those below 35.

TABLE IX.—PER CENT RECRUITS ENTERING HIGH SCHOOL, COLLEGE, ETC., AT VARIOUS ALPHA LEVELS

Alpha score	School continuance		
	Per cent entering high school	Per cent high school seniors	Per cent entering college
155 and above	93	73.0	53.0
135–154	84	55.0	39.0
115–134	72	38.0	21.0
95–114	55	22.0	11.0
75– 94	45	14.0	7.0
55– 74	23	6.0	3.0
35– 54	10	2.0	1.2
Below 35	4	0.7	0.3
Total	36	22.0	20.0

Figure 3, illustrating the second column of Table IX, shows similarly the proportions who reported that they became seniors and (practically all of them) graduated from high school. Of those scoring less than 35, less than 1 per cent, and of those scoring 155 or over, 73 per cent, reached the senior year in high school. Thus the chance of reaching this level is over 100 times as great for the highest as for the lowest group. Here again the contrast would be intensified had we had a comparable measure of the illiterate group, and included them.

Figure 4 shows comparable figures for entrance to college. A quarter of 1 per cent of the lowest group, and 53 per cent of the highest group, reported that they had entered college. The chance

of college entrance at that time appears to have been almost 200 times as good for those highly endowed intellectually as for the lowest fifth. Were illiterates included, the contrast would in this case also be strengthened.

This educational selection of intelligence is evidenced also in the large yearly increase in Alpha medians when the test is given throughout a school. Of course not the whole of this increase is due to elimination of the less intelligent pupils. To determine the exact amount which is due to this cause is at present impossible, but its existence is easily proved. Even the amount by which Alpha scores increase each year is not readily determined from data so sketchy as are these. The average figures from Table X are, from freshman to sophomore year, 15 points; from sophomore to junior year, 10 points; and from junior to senior year, 4 points. Fifteen points is approximately the amount by which, in the lower part of the scale, the score increases with an increase in *mental* age of 1 year. If 15 points up here is equivalent to 15 points lower down on the scale (which is quite problematical) then high school pupils, after the first year, are growing more slowly mentally than when they were younger.

The total increase from freshman to senior year may be estimated at about 30 points. Tables X and XI facilitate a comparison of Army figures with school figures, and an estimate of those elements which in the Army group were constant. Maturity, for instance, plays no part in the difference in the Army figures, for the men were all examined at the same time, after manhood was reached, instead of at different stages in the growth period; the separation into groups those who had left school as freshmen, sophomores, etc.—was made afterwards. Part of the effect of instruction also is not present in these Army differences—whatever part is temporary, and is afterwards forgotten and lost. But the groups of recruits do differ, and this remaining difference must be due to the effect of educational selection, and the more permanent effects of instruction. This remaining effect is strikingly less than the immediate effect of growth and instruction which we find in the groups examined year by year while in school. In general, it is not much over half as great. In other words, the indication is that mental maturity—mere mental growth, independent of environment—together with the temporary effects of instruction, is responsible for a good half of the change of score from freshman to sophomore year; from junior to senior year it accounts for about one-fifth of the change. Educational selection, and the permanent effects

of instruction, seem to account for nearly half of the increase from the first to the second year, and almost the whole of the yearly increase later on. Improvement in native intelligence, *i.e.*, inner development, or mental growth apart from instruction, is almost certainly still going on when these children enter high school. It is almost equally certain that very little of it goes on during their last year in high school.

TABLE X.—HIGH SCHOOL ALPHA SCORES

	Schools						Army	
	Mich.	Ill. Wis. and Ia.	N. Y. (Kan. report)	Emporia, Kan.	Stanton, Va.	Minn. survey	Recruits	Officers
Freshmen.....	88	99	92	80	91.0	93	98	140.7
Sophomores...	104	112	104	105	114.5	105	105	141.3
Juniors.......	118	124	118	101	136.0	111	111	142.0
Seniors.......	123	127	132	111	117.0	120	115	142.6

TABLE XI.—YEARLY INCREMENTS

	Schools						Army	
	Mich.	Ill. Wis. and Ia.	N. Y. (Kan. report)	Emporia, Kan.	Stanton, Va.	Minn. survey	Recruits	Officers
Freshman to sophomore..	16	13	12	25	21.5	12	7	0.7
Sophomore to junior......	14	12	14	−4	21.5	6	6	0.7
Junior to senior.........	5	3	14	10	−19.0	9	4	0.6
Freshman to senior......	35	28	40	31	24.0	27	17	2.0

In the officer group, maturity not only, but also selection and differences in amount of instruction are eliminated from the question, since very few men got into the officer group who had not had college training. Practically all of them are present in the group for each high school year, the drop in numbers being less than 3, 5 and 5 per cent for the three intervals. Accordingly, with none of the causes of

difference present, we should expect to find almost no difference in the Alpha scores. Actually, for each year interval the increase is less than one point Alpha score.

The influence of intelligence on continuance in school appears also when we look at the elimination which takes place among the freshmen of least intelligence. Since the facts from our previous tables indicate that a freshman who scores 77.5 will as a senior score about 90, we can get a rough notion of the extent of this elimination without necessarily following a freshman group all the way through. The assumption must be made that successive entering classes are of approximately the same size; then, we may compare the number in the freshman class who score below 77.5 with the number in the senior class who score below 90. Comparisons of this kind show that in the Michigan schools, 87 per cent of the freshmen below 77.5 drop out before senior year. In Madsen's group (Ill., Ia., and Wis.) about 84 per cent drop out. In general, at this Alpha level of intelligence, only about one in seven remains to graduate.

In the Army group, no adjustment between freshmen and senior scores needs to be made, since (as explained above) all the tests were made several years after the men left school. Here we find, among the recruits, that 78 per cent of those below 85 Alpha were eliminated. Among the officers, where the total elimination amounted to only about 10 per cent, 24 per cent of those scoring below 85 were nevertheless eliminated before senior year.

We have said that children who at 14 years of age score less than 60 to 65 on the Alpha examination are not likely to enter high school. Can we now, in summary, make a similar statement about the probability of graduation? Table VII shows that in the best high schools very few seniors score below 90 (MA about 15). We may estimate that this corresponds to a freshman score of 77.5 and can then say that not more than 1 freshman in 6 or 7 of those who as freshmen score below 77.5 (MA 14 years) remains to graduate.

(*To be Concluded in December.*)

THE PROBLEM OF GROUP INTELLIGENCE TESTS FOR VERY YOUNG CHILDREN[1]

R. PINTNER

AND

BESS V. CUNNINGHAM

Teachers College, Columbia University

The practical value of group intelligence tests for the better classification of school children is so great that a large number of such tests have appeared during the last 5 years. The easiest to construct and, therefore, the most numerous are the so-called verbal intelligence tests which involve a knowledge of reading and writing. These cannot be given to much advantage below Grade II. It is, however, precisely in Grade I and kindergarten that intelligence tests are of supreme importance, because in these grades the teachers have few or no measures of the children in school attainment upon which to base an estimate of their ability. Furthermore, it is of great practical value to group children as early as possible in relatively homogeneous groups, so that they may start their school career happily and properly adjusted.

A valuable list of the available group tests has been published by Whipple in "The Twenty-first Yearbook of the National Society for the Study of Education." In this book also there appears a good discussion of tests for the lower primary grades by Rogers. A comparative study of four of such tests has been made by Henmon and Streitz.[2] These authors conclude that there are no striking differences between three of the scales, and that while none of the three is a perfect measuring instrument, nevertheless each of them contributes valuable information as to the intelligence of the children measured.

It has been maintained by some that it is not possible, or at least not feasible, to test kindergarten children by means of the group method. They are supposed to be so independent and individualistic,

---

[1] The tests described in this article are published by the World Book Company, Yonkers, N. Y., and are called "The Pintner-Cunningham Primary Mental Tests."

[2] Henmon, V. A. C., and Streitz, R.: A Comparative Study of Four Group Scales for the Primary Grades. *Journal of Educational Research,* Vol. V, No. 3, March, 1922, pp. 185–194.

so little amenable to group control, as to make impossible the giving of a group test. It is certainly true that children at this level are independent and individualistic and, in addition, little habituated to follow directions and commands given to them as a group. The more modern or less formal the kindergarten is, the less practiced are they in group work. This does not mean, however, that group tests are, therefore, impossible. It means rather that the psychologist must make his tests of such intrinsic interest to the child as to compel his attention, so to devise them as to make the child feel that he is playing a very special game along with his fellows. Under such conditions the size of the group that can be handled by one examiner will depend upon the skill of the examiner in dealing with young children and upon the number of assistants that may be available. It has been the experience of the writers to find it perfectly feasible for one examiner without assistance to test as many as 25 kindergarten children at one time. It may not be desirable in general to do this, and it may be good policy in most cases to restrict the group to about fifteen children.

In the construction of a suitable kindergarten and Grade I test, there are certain necessary prerequisites. The test should contain no letters or numbers. Although some children in Grade I and even in the kindergarten are familiar with numbers and with some letters or even words, the vast majority are not; and the introduction of such material will tend to convert the test into an achievement test rather than a test of native ability. Furthermore, the responses required of the child must not involve the writing of conventional signs, such as letters or figures. Only the simplest kind of response with a pencil or crayon should be demanded and, if possible, this response should be uniform, or nearly so, throughout the test. In the tests we have constructed the child responds by marking something and the mark in each case is a simple line drawn on a picture or element of the test. The one exception to this is the dot drawing test in which the child has to draw a line from one dot to another as in the copy before him.

In 1920, a first set of tests was constructed consisting of five exercises: (1) Recognition of common objects used in various situations; (2) the finding of isolated parts of a picture; (3) the connecting of dots so as to copy a given simple picture; (4) checking the pictures described in a story told to the children. These tests were given to about one hundred children ranging in age from 4½ to 7 years. A careful analysis of the results of each element in each exercise was made.

In addition correlations with the Binet, Dearborn, Detroit, Kingsbury and Pressey were computed. The coefficients ranged from 72 with the Binet to 49 with the Detroit.

Four more exercises were then constructed as follows: (1) æsthetic differences, *i.e.*, marking the prettiest of three similar objects; (2) marking two associated objects in a series of four objects; (3) drawing completion; (4) marking the shortest distance between two given points, a sort of simplified maze test. A minute analysis of these various tests and the elements of each was again made. As a result of this, several changes were made and a few of the tests omitted. The tests were then tried out on six first grades and a further analysis of the results made.

On the basis of this experience a preliminary edition of the tests was printed. This First Revision consisted of a booklet of eight pages measuring $8\frac{1}{2} \times 11$ inches, that is, the conventional size for test blanks. There were six exercises as follows:

Page 2—Common Observation—5 elements;
Page 3—Æsthetic Differences—6 elements;
Pages 4 and 5—Picture Parts—8 elements;
Page 6—Associated Objects—7 elements;
Page 7—Picture Completion—12 elements;
Page 8—Dot Drawing—12 elements.

In each exercise the elements progressed from easy to harder ones. Rather extensive trials with this edition showed that the test discriminated well between ages 5, 6 and 7, and fairly well between the half age intervals. The correlation of the test scores with Binet mental ages of 18 cases, ranging in chronological age from 5-1 to 7-11 was 0.87. The correlations of the same cases with each of the separate tests of the group ranged from 0.60 to 0.76. A correlation of 27 cases of superior 7- and 8-year-olds, who were in Grades III and IV, between the Binet and the test scores showed a correlation of only 0.48, revealing a deficiency in discriminating capacity for brighter children. The rank correlation between Binet and test score of 33 children, mostly foreign, in a special school, ranging in MA from 7 to 9-6 and in IQ from 46 to 87, was 0.46.

A number of cases tested on the Binet Scale showed the following average scores for the mental ages of feebleminded and bright children:

Mental age	Feebleminded		Bright	
	n	Score	*n*	Score
8-0–8-5	14	54	5	54
7-6–7-11	16	47	3	46
7-0–7-5	18	36	5	42
6-5–6-11	9	31	3	28

On the basis of this experience with the test certain changes were now made. The large page of the test blank measuring 8½ by 11 inches was abandoned and a small page measuring 6 by 9½ inches was substituted. This change was felt to be particularly desirable for the kindergarten children. Through the cooperation of Professor Patty Hill of Teachers College, Columbia, the youngest kindergarten children of the Horace Mann School were examined individually. Actual experience in watching these younger children perform the test individually, listening to their remarks and questioning them at times showed a marked tendency for them to be distracted by the relatively large number of pictures on the large page.

They found great difficulty in keeping their attention on the item under consideration at any one time.[1] Several items of about the same difficulty were omitted and in their places a few harder and easier items substituted. The picture completion test was so constructed as to avoid the necessity for the child to draw in the missing part. This was accomplished by presenting several parts near the incomplete picture, one of which had to be marked. A new test, called discrimination of size, was added. This presents a doll and the child has to choose from three dresses, hats, shoes, gloves, the one that will best fit the doll. In addition several minor changes in arrangement and spacing were made in order to help the young child keep his attention on the item under consideration.

In the final edition of the test, therefore, we have a booklet of 16 pages measuring 6 by 9½ inches. The first page is for the name, age and other necessary data as well as for a record of the scores. No

[1] Cf. Rogers, A. H.: Measurement of the Abilities and Achievements of Children in the Lower Primary Grades. *The Twenty-first Yearbook of the National Society for the Study of Education*, 1922, pp. 143–151.

page contains more than four items of a test and several contain only one or two. In this form it has been given to about one thousand children, between the ages of 4 to 8, and has been found to work very well. The problem of turning over the pages is greater in the new form than in the old, but the children quickly become adapted to it with a little help and guidance. With the small page the common tendency of the young child to be distracted by the numerous pictures is very much reduced. Time limits for each exercise have been set, not with the idea of making the test a speed test, but to insure that it shall be given to all alike under standard conditions.

Some of the correlations of this test with other tests are as follows:

Test	Cases	Grade	Correlation
Binet	19	Grade I children	0.82
Binet	20	Kindergarten children	0.71
Binet	17	Kindergarten children	0.55
Binet	105	Kindergarten and Grade I children	0.77
Kingsbury	74	Grade I and Grade II children	0.56
Otis Primary	39	Grade II children	0.66
Otis Primary	36	Kindergarten children	0.66
Teachers' ranking	19	Grade I children	0.78
Teachers' ranking	26	Grade II children	0.64
Teachers' ranking	36	Kindergarten children	0.78

In a survey of two schools under the direction of Dr. McCall of Teachers College this test was used. A composite rating based upon the teachers' rating—the Otis in one case, the Kingsbury in the other, and our own test—was devised. This composite rating was made for practical purposes in the survey and not for a measure of our test. The following are correlations of our test with this composite:

Grade II = 0.78 .... 26 cases
Grade II = 0.79 .... 39 cases
Grade I = 0.81 .... 44 cases
Kindergarten = 0.83 .... 36 cases

With two groups of children the test was repeated after an interval of one day to measure its reliability. The correlations of the first with the second trial are:

17 Kindergarten children 0.88
20 Kindergarten children 0.96

A distribution of the scores for each half age and for each age from 4 to 8 has been made, and percentile scores computed. The median scores are shown below:

For Half Year Intervals

Age	Median	N
4-0–4-5	10	7
4-6–4-11	13	34
5-0–5-5	15	79
5-6–5-11	18	148
6-0–6-5	24	234
6-6–6-11	28	212
7-0–7-5	34	176
7-6–7-11	39	92
8-0–8-5	42	50
8-6–8-11	38	23
Total		1055

For Whole Year Intervals

Age	Median	N
4	12	41
5	18	227
6	26	446
7	36	268
8	40	73
Total		1055

Figure 1 shows the percentile curves for ages 4, 5, 6, 7 and 8. The actual percentile points calculated are shown on the base line. These curves show a discrimination at all percentile points except in two cases at the 0 and 100 points. Our sampling of age 8 is probably not so good as the samplings at the other ages. Similar percentile curves for the half ages have been constructed. These naturally do not show such a good discrimination as for the whole ages.

A percentage distribution of the scores according to half ages is given in Table I, and according to whole ages in Table II. Frequency curves of these distributions have been constructed but are not reproduced here, since the main facts can be gathered from the tables and the percentile curves, and they do not add anything of material importance.

Our norms are by no means adequate at the present time. The median for age four is in all probability too high, and the median for age eight is probably rather low. These norms, therefore, must be

regarded as merely suggestive. In the near future it is hoped that more adequate norms will be available. In the meantime, however, the test may be of use in classification of children. The correlations with other tests and with teachers' ratings are on the whole fairly high. The test should prove of decided value in practical work for the classification of kindergarten and Grade I children.

The practical value of the test was demonstrated by an experiment in classification of Grade I children in Newton School, Toledo, Ohio. The Grade I children were all tested at the beginning of the school year by means of the first printed edition of the Pintner-Cunningham

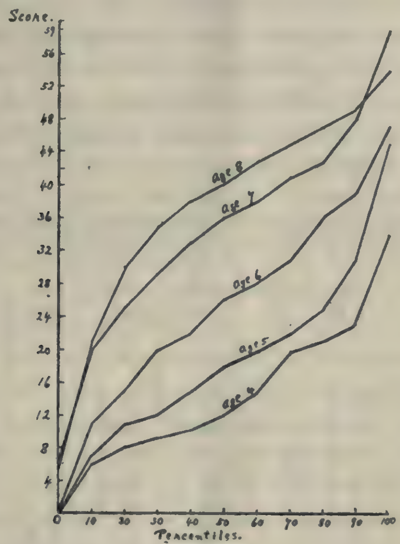


FIG. 1.—Percentile curves for ages 4 to 8.

Tests. Upon the basis of test results the children were placed in three rooms—the ones making high scores upon the tests in one room, those making average scores in another and those making poor scores in still another. The complete results of the study can be determined only by the use of further tests, but the regular reports of principal and teachers indicate that the experiment has been highly successful. There have been no failures in promotion in either the bright or the average groups. Bright and slow children have profited equally by the division of the classes into groups as nearly homogeneous as possible. The bright children have covered 2 years work in 1 year in some cases. While the advisability of rapid advance as an aim

is not fully established, the results at least prove the need for devising some means of better meeting individual needs in the schoolroom. Children are happier when working neither too far beyond nor too far under their maximum capacities. The use of group mental tests can at least help in classification.

The mental test measures only one phase of activity. Within a year a companion series to the mental test is to be published. The aim of the combined series is to test achievement in responses which may be taught upon a basis of native capacity to learn. Beginning reading, number work, drawing, and a knowledge of right social relations are to be included in the series.

Group tests can best serve the kindergarten and primary child and teacher if used sympathetically, with a full appreciation of the fact that they aim only to help in the better understanding of the needs of each child and are to be considered as guides rather than as conclusive evidence of final procedure.

Table I.—Percentage Distribution of Scores for Half-age Intervals

Scores	4-0–4-5	4-6–4-11	5-0–5-5	5-6–5-11	6-0–6-5	6-6–6-11	7-0–7-5	7-6–7-11	8-0–8-5	8-6–8-11
0–9	28.6	29.4	16.5	12.2	13.5	3.8	....	1.1	2.0	4.3
10–19	42.8	38.2	45.5	43.3	21.4	19.8	7.4	6.5	6.0	4.3
20–29	28.6	29.4	31.6	32.5	35.1	34.9	26.2	15.2	10.0	13.0
30–39	....	2.9	5.1	11.5	25.6	26.4	34.7	30.4	28.0	34.8
40–49	....	....	1.3	0.7	4.7	15.2	29.0	39.1	46.0	43.5
50–59	....	....	....	....	....	....	2.8	7.6	8.0	

Table II.—Percentage Distribution of Scores for Year Intervals

Scores	4	5	6	7	8
0–9	29.3	13.6	8.7	0.4	2.7
10–19	39.0	44.0	20.6	7.1	5.5
20–29	29.3	32.0	34.9	22.4	11.0
30–39	2.4	9.3	25.9	33.2	30.2
40–49	....	0.9	9.6	32.6	45.2
50–59	....	....	....	4.5	5.5

girl was given especial care and careful training after the initial warning to the parents.

There is another factor which introduces complications into this case. At the time of the first examination the girl was reported to be physically retarded. Comparisons of her height and weight with Baldwin's Standards show that in 1918 she was 0.593 AD below children of her age in height, and 1.31 AD under the average weight. In 1920 she was 0.032 AD *over* height, and only 0.468 AD under weight. A study of the ossification of her wrist bones and of her dentition at this later date indicated that she was at least normal, and possibly somewhat accelerated. All these facts indicate that there has been at some time during the 4 years a period of rapid physical growth, and so it is not surprising to find in the changes of the Intelligence Quotient some evidence of a rapid mental development as well.

A most interesting case is presented by No. 16, a girl who was 9 years 11 months when first tested. She secured a mental age of only a month less, which made her IQ 99. There was nothing out of the ordinary about her examination except that her reactions were reported very slow. The teachers reported that she was quiet, reserved, bashful, timid, repressed, and slow in motor reactions. At the end of the summer, however, all reports noted that she had improved to a marked degree. As one teacher put it, she was "slowly coming out of her shell."

The next year, when she was 10-11 she earned a mental age of 12-3, and her Intelligence Quotient rose to 112. The examiner reported as follows: "Her IQ may be higher. She has great trouble in expressing herself. Standards of her own answers are very high. Won't answer unless she is sure; refuses to guess." The teachers made the same kind of reports this year.

On the next examination, taken when she was exactly 12, her mental age was 14-2. This raised her Intelligence Quotient to 118. Once more the teachers reported her as slow, reticent, bashful, repressed, at the beginning of the summer, but noted later in the summer very marked changes. One report says: "Remarkable development during the last week."

In 1921, when she was 13, her mental age was 17-8. This makes her Intelligence Quotient 136, a gain of 37 points since the first examination. The teachers reported her as quiet and reserved, but not so much is made of this side of her nature as in the earlier years. The physical training instructor makes a very significant statement: "She has passed through a period of very rapid growth."

It is in this rapid growth that we may find the explanation of the rapid mental development. The contrast is well shown when her height and weight are compared with Baldwin's standards. Her

	1918	1921
Height..............	0.586 AD	1.40 AD
Weight..............	0.325 AD	0.511 AD

physical development was retarded at the time of the first examination and it is reasonable to infer that there was mental retardation as well. With the acceleration of the one came the acceleration of the other.

Number 22 was first examined when she was only 2 years 8 months old. Her performance was somewhat out of the ordinary, in that she was able to obtain credit for the description of pictures in year VII, in spite of the fact that she missed all the tests in year V and all but the alternate in year VI. As she had been given no opportunity to learn the names of coins she was given the time orientation test in VI and passed. With these credits her IQ is 138, without them it drops to 125.

She was not examined again until 2 years later, when she was 4 years 8 months old. This year she was not able to describe the pictures, though she had done it creditably 2 years before. No allowance was made for her lack of acquaintance with the coins, and her IQ on this basis was 125. Had she obtained credit for picture description and the time orientation in VI, the resulting IQ would have been 133.

On her last examination (1922) at the age of 5 years 9 months she made a mental age of 6 years 10 months, with an IQ of 119. She did not know the coins at this time, and credit for the time orientation was not given. With this credit the IQ would have been 122.

This case presents a very good illustration of a large difference in IQ resulting from what appears to be a chance success, for it seems that we must attribute to chance the successful description of pictures by a subject only slightly over 2½ years old. We also find a complication due to the use of an alternate test, and the question arises as to how far this is permissible. Undoubtedly it was justified in the first examination, but the case is not so clear when the subject is nearly 6 years old.

Case No. 23, a girl, was 9 years old when first examined, but tested only 7-8, which gave her an IQ of 85. She missed the coins in VI, could not draw a satisfactory diamond in VII, and failed on three tests in VIII; Ball and Field, counting backwards, and definitions.

In two subsequent years her Intelligence Quotients were 102 and 100. The explanation of this rise may be found in the fact that this girl was confined to her bed during the first 5 years of her life, and so did not have the range of experiences of the ordinary child. The first test came before she had been out in the world long enough to catch up in her development. Another factor has a slight influence in this case. In each of the later years there was one test on which she was given credit for rather doubtful responses. Giving her the benefit of the doubt raised the IQ's two or three points. Also, on the first examination there were two tests which she barely missed. Had she passed these her IQ would have been at least four points higher.

Case No. 24 is a boy who was 8 years 6 months old when the first examination was given. He tested 9-8, which made his IQ 114. The examiner reported that he was seemingly only partly interested in the work, and this seems to be borne out by the fact that he scattered widely, as the basal year was VII, and he passed tests in every year to XVI, inclusive. A little later he was examined on the Doll Short Test, and secured an IQ of 125. The teachers reported that he was prone to bluff, was self-centered and superficial.

A little over a year later he was 9 years 8 months old, and tested 10-10 with an IQ of 112. He failed many of the tests in which he had been successful the previous year, and but for this would have had a much higher IQ, as these misses total nearly a year. It is also noticeable that he barely misses several tests. For instance, he interpreted one picture, and gave partially interpretative responses for two others. Also, he missed the arithmetic problems in XIV because he gave the answer to the first as 50 days rather than 50 weeks. On the Yerkes Point Scale examination he secured a CIA of 124.

No examination was given to this subject in the following year, but in 1921, when he was 11 years 8 months old, he tested 14-6 with an IQ of 124. It is interesting to note, in view of the wide scatter on the first examination, that on the last the base was XIV, and only one other test was passed—Binet's Paper Cutting Test in XVIII.

It seems reasonable to suppose that the lower IQ's resulting from the first examinations were due to the failure of the subject to put forth his full effort during the testing. The fact that each year he did better on other examinations is evidence in support of this supposition. This case offers a very convincing argument for the scientific method of determining an IQ on the basis of a number of measurements.

Case No. 25 presents what is perhaps the hardest problem for

analysis in the whole group. At the time of the first examination he was 10 years 5 months old, and he tested 12 years 11 months with an IQ of 124. The next summer he was 11-7 when examined, and he tested 15-2 which gave an IQ of 131. The third examination found him 12 years 6 months old, and he earned a mental age of 17-0 by getting all the tests in XVI and two in XVIII. This gave him an IQ of 136, 12 points above that obtained on the first examination.

Unfortunately, we have not the data to check up this boy on the physical side, as his height and weight were not reported for the third year. It is possible that some explanation might be found in accelerated growth, but as he was well above the average in 1918 this does not seem likely. The writer is inclined to believe that here we have differences due to the chance success or failure in tests of the upper years where the credit for a single test is large. If he had passed one test from year XVI on his first examination his IQ would have been raised four points. On the other hand, if he had missed one of the year XVIII tests which he got on the last examination there would have been a drop of four points in the IQ.

Number 26 is the brother of 25. He was first examined when he was 6 years 1 month old and tested 8-11. This gave him an IQ of 147. He had been sickly since birth, and was reported as very unevenly developed. The teachers reported that the summer's work made a great improvement in every way.

On the second examination when he was 7-3, he tested 10-5, with an IQ of 144, and he got 7 points out of the 8 necessary for credit in the fable test of year XVI. Had he passed this test his IQ would have been 150.

On the third and fourth examinations he made IQ's of 166 and 167, a gain of about 20 points over the first two. It seems likely that the results of these later examinations show his correct mental status. At the time of the first and second examinations he had not been taught to read, and because of this he did not have the necessary background for answering correctly some of the more advanced tests. Also, his poor health had prevented the acquiring of much social experience, as he was not in school with other children. With his improving health and his entrance into school these deficiencies were made up, and he was able to reach his true level on the later examinations.

The cases to be considered in the second part of this study (See Table II) are school children examined by the staff and students in

training at the Psycho-Educational Clinic of the Harvard Graduate School of Education. For the most part these were children who were not getting along in their school work, and the examinations were made in the attempt to determine how the methods of their education could best be altered to suit their needs and abilities. The time between examinations ranged from less than a month to 15 months.

The distribution of the differences is as follows:

Points	0	1	2	3	4	5	6	7	8	13	19
Number	3	2	3	1	2	2	0	1	1	1	1

The median difference is 3.5, and only four cases show a difference of more than 5 points.

TABLE II.—INTELLIGENCE QUOTIENTS OF CLINIC CASES

Number	Date	First test		IQ	Date	Second test		IQ
		CA	MA			CA	MA	
5	May, 1920......	12–8	10–4	81	Oct., 1920......	13–2	11–5	89
38	Jan., 1920......	8–0	7–4	92	May, 1921......	9–4	7–11	85
44	Apr., 1920......	12–8	9–9½	77	Oct., 1920......	13–2	9–5	72
45	May, 1920......	12–2	8–4	69	Oct., 1920......	12–6	8–8	69
49	Jan., 1920......	5–9	3–10	67	June, 1921......	7–2	6–2	86
59	Jan., 1920......	9–6	7–9	83	May, 1921......	10–11	8–10	81
67	Apr., 1919......	7–9	9–0	116	May, 1920......	8–10	11–1	113
101	Nov., 1919......	8–1	6–6	80	Dec., 1919......	8–1	6–6	80
111	Oct., 1919......	8–5	7–6	89	Feb., 1921......	9–9	8–2	84
121	Apr., 1920......	13–0	8–8	66	Oct., 1920......	13–7	9–4	68
123	Jan., 1920......	5–11	4–10	82	Nov., 1920......	6–9	5–10	82
130	Jan., 1920......	5–6	4–2	76	Nov., 1920......	6–5	5–9	89
133	Nov., 1920......	12–11	16–5	127	Jan., 1922......	14–1	17–4	123
147	Apr., 1919......	11–8	13–10	119	May, 1920......	12–9	15–6	121
148	Jan., 1920......	9–2	7–4	80	May, 1921......	10–6	8–0	76
171	Dec., 1919......	9–9	8–4	85	Nov., 1920......	10–9	9–3	86

There are only two cases in this group which need discussion. Number 49 was first tested at the age of 5 years 9 months when he was in Grade I. He earned a mental age of only 3-10, which made his IQ 69. It was found upon inquiry from the school authorities that he had spent the whole previous year in the kindergarten without speaking once. In every other way he seemed normal. His parents were recently arrived immigrants. In the form board tests he did much

better than his mental age would seem to warrant, and diagnosis was deferred with the recommendation that he be retested later. On the reexamination, which was given 15 months later, he made an IQ of 86. It seems reasonable to suppose that the difficulties which he experienced on the first examination were largely linguistic. As he gained a greater command of English he was able to pass more of the tests, and so made a higher mental age. This supposition is borne out by the fact that most of his failures on the first examination are found to be in the linguistic tests.

Number 130 was first tested when he was 5 years 6 months old. His mental age on the examination was 4-2, giving him an IQ of 76. Nine months later he obtained an IQ of 89. Investigation showed that he had only been in school a few months at the time of the first examination, and that he had not been in the same school during that time. He did not even know his name when first tested. He came from a markedly inferior home, and it seems likely that the conditions of his environment were such that he did not have the opportunity to learn as the normal child of his age. In the school some of this deficiency was remedied, and thus the improvement on the second examination.

The causes of the Intelligence Quotient differences which have been pointed out in the previous case studies may be grouped under a few general heads. We find in the first place indications that some differences are due merely to fluctuations in ability, interest, attention, or whatever we may choose to call it. This is a very common phenomenon in all phases of endeavor, and it would be very remarkable if it did not appear in the results of mental examinations. It is probable that the differences which may be attributed to this cause are for the most part small, but at times the fluctuations may be large and the IQ differences great. These fluctuations are more significant than they otherwise might be on account of the large units in which the Binet Scales measure. It seems likely that many of these differences would disappear if we adopted the more scientific plan of taking several measurements instead of one.

In some cases the fluctuations are so great that they can not be said to indicate simply temporary variations in efficiency but rather an unstable or psychopathic personality. It seems that this conclusion must be reached when a subject fails in a considerable number of tests which he passed in previous years. This factor has been present in at least two of the cases presented in this study, as shown by the supplementary information obtained.

There are undoubtedly children who are temporarily retarded in their mental growth as they are retarded physically and physiologically. In these cases it is reasonable to expect a substantial increase in the Intelligence Quotient as the development proceeds.

Sometimes the differences may be due to some deficiency in the training or environment of the subject. If a child has not had the ordinary education and experience of the average child of his age he can not be expected to pass all the tests. Differences of this sort may be expected to appear when the subject has not become thoroughly familiar with English or when poor health has kept him from school or from the usual play activities in which he might be expected to take part. This factor is also of considerable influence in the case of a very superior young child, as he has not had the experience which would enable him to pass the more advanced tests.

The record which any individual makes on an intelligence examination is not due to his native capacity or natural endowments alone. It is influenced by his mental health, the stage of development of his innate ability, and the general environment, including formal and informal education, to which he has been subjected. Most children are mentally sound and sane, they develop at nearly the average rate, and they grow under the ordinary circumstances of environment. In such cases substantial constancy of Intelligence Quotients may reasonably be expected. This study seems to indicate pretty clearly that when large differences in Intelligence Quotients appear there may be found some reason for them. Thus the discrepancies do not lessen the value of the Intelligence Quotient but increase its usefulness, providing always that a scientific case study of the individual is made.

A REPORT ON THE CORRELATION OF PSYCHOLOGICAL TESTS WITH ACADEMIC AND MANUAL SUBJECTS [1]

IRENE GLENN

North Bennett Industrial School, Boston, Mass.

The work reported on in the following was undertaken on the hypothesis that individual differences in vocational fitness exist among children of the ages found in the grades preceding high school, and that they exist to such an extent that all children cannot efficiently be included in the same curriculum. We have sought to determine where some of these differences lie and by what tests they may best be indicated.

The subjects for this experiment were in the Grades VI, VII and VIII. It is in Grade VII in Somerville that manual training is introduced in the curriculum, Grade VI including only the simplest hand work and drawing. In general the children came from lower middle class homes and represented a variety of races.

The following tests were used:

General Intelligence

Binet-Simon (Terman Revision, Abbreviated Scale).
Pintner Non-language Mental Test.
Myers' Mental Measure.

Academic

Woody-McCall Mixed Fundamentals, Form II.
Trabue Language Scale C.
Thorndike Scale Alpha 2 (Parts 1 and 2) for Measuring the Understanding of Sentences.
Directions Test (hard).

Motor Tests

Healy Psychomotor (tapping) Test (used as a group test).
Star Test (individual).
Wells' Peg Test (individual).
Paper-folding Test (individual).
Lane Test (individual).

---

[1] This article is a report on a part of the research which is being supported and conducted by the North Bennet Street Industrial School, Boston, Mass. The School is indebted to Mr. Charles S. Clark, Superintendent of Schools, Somerville, Mass., for his cooperation in opening his classes to the experimentation.

Except for the directions test, the academic tests are fairly analogous to the academic subjects in the curriculum. In choosing the motor tests, our main object (aside from simplicity and practicability) was that they should be as purely motor as possible. Puzzles or construction tests of a problem nature, though involving no language, have been avoided. The work was originally planned so that all classes were to receive all the tests, but the addition of further tests and the time required by others, operated to limit the number of classes to which certain tests could be given. All examinations for a given test came as near as possible in point of time. For each of the standard tests, the procedure and scoring used was that given by its author. The method of scoring for the Thorndike Alpha 2 was that adapted to individual testing by T. L. Kelley.

Detailed instructions for ranking pupils on class work were given to each teacher, the method being that of selecting the best, the poorest, and an average child, then selecting the children whose ability was between the extremes and the average, and finally filling in all others in relation to these five. The coefficients were calculated for each class separately. The footrule formula was used, R being transmuted into r.

On the whole the correlations are low but this is to be expected since the groups are small and fairly homogeneous. Academic subjects show a marked tendency toward a negative relation with motor

TABLE I.—SHOWING NUMBER OF MEASURES USED IN CALCULATION OF COEFFICIENTS

Grade	Class	Total number of children (*i.e.*, number taking academic subjects)	Number of boys taking manual training	Number of girls taking manual training
VI	VI3	22		
	VI1	45		
	VI2	37		
VII	I1	42	25	17
	I2	46	28	18
	I*F*	34	22	12
VIII	II2	40	29	11
	II3	31	. .	19
	II*F*	35	14	

TABLE II.—AVERAGES OF COEFFICIENTS OF CORRELATION FOR GENERAL INTELLIGENCE AND ACADEMIC TESTS WITH SCHOOL SUBJECTS[1]

	Pintner	Myers	Binet	Woody	Trabue	Thorndike	Directions
English	.20(8) .04– .46	.27(5) –.11– .35	.28(5) –.11– .46	.41(8) .24– .67	.34(8) –.06– .62	.59(3) .55– .65	.69
History	.32(6) .11– .55	.38(4) .28– .54	.56(4) .47– .61	.29(6) .04– .57	.32(6) .19– .59	.41(2) .41 & .41	.43
Geography	.17(2) .04 & .30	.16(2) .09 & .22	.40(2) .32 & .49	.60	.37	.52	.71
Arithmetic	.27(9) .01– .50	.26(6) .11– .34	.28(6) –.11– .55	.67(8) .47– .86	.23(8) .04– .47	.48(4) .28– .80	.38(2) .29 & .47
Woodworking	.35(5) .12– .88	.19(3) –.23– .46	.14(3) .00– .31	.11(5) .04– .26	–.03(5) –.42– .33	.06(2) –.08 & .19	.42
Woodworking	.55(3) .37– .90	.14(2) .08 & .21	.30(2) .23 & .37	.38(3) .34– .46	.12(3) –.12– .50	.23(2) .01 & .45	–.11
Sewing	.14(5) –.09– .39	.04(4) –.14– .38	.24(4) –.18– .73	.28(5) .13– .43	.08(5) –.13– .26	.18(2) .00 & .37	–.14
Cooking	.26(4) .12– .52	.32(3) –.02– .50	.31(3) .01– .55	.07(4) –.34– .34	.12(4) –.26– .82	.30(2) .08 & .52	.00
Bookbinding	.25	.31	.02	–.21	.09		

tests. On the other hand their correlations with the academic tests are on the whole higher even than those with general intelligence. That is, of course to be expected. The relation of the manual subjects to the tests is irregular, but there are very definite tendencies toward the relation being positive in the case of the manual tests and negative or very low in the case of the academic tests. The greater variability evident here, may be due, in part, to other causes than factors inherent in the tests themselves. The criteria whereby a teacher may rank pupils in manual work are not so exact as those for academic work. Character traits of confidence and obedience, habits of industry and neatness, are all more evident in classes of manual training where groups are smaller and freedom greater. In consequence, we can not expect that the estimates given by the teachers for these subjects shall represent only the manual skill of the pupils. The tests, on the other hand, exclude to a great extent everything except manual skill.

[1] The figure in parenthesis indicates the number of coefficients represented in the average and beneath is given the range of the distribution of the coefficients.

TABLE III.—AVERAGES OF COEFFICIENTS OF CORRELATION FOR MOTOR TESTS WITH SCHOOL SUBJECTS

	Psychomotor	Wells	Star	Folding	Lane
English	.01(8) −.18−.18	−.13(5) −.41−.17	−.05(3) −.34−.17	−.02(3) −.10−.03	.03(3) −.05−.11
History	−.10(6) −.28−.18	−.09(4) −.32−.11	−.04(4) −.21 & .14	−.03(2) −.13 & .19	−.06(2) −.18 & .06
Geography	−.14(2) −.26 & −.02	.24(2) .12 & .37	.05(2) −.02 & .12		
Arithmetic	.00(9) −.20−.26	.04(6) −.32−.34	−.10(4) −.43−.15	.11(3) −.10−.40	−.05(3) −.13−.00
Woodworking	.22(5) −.12−.43	.13(3) −.05−.22	.22(2) .21 & .22	.50	.21(2) .21 & .21
Woodworking	.29(4) .09−.50	.14	....	.49	.34(2) .30 & .38
Sewing	.30(5) .08−.58	.28(3) −.18−.68	.11	.67	−.03(2) −.05 & −.01
Cooking	.04(4) −.18−.40	.08(2) −.26 & .41	.11	.08	−.24(2) −.37 & −.12
Bookbinding	.43	.24	.41		

We find, then, that our hypothesis regarding the differentiation of abilities at the ages represented by our subjects has a basis in test results as well as in the experience of educators. Speaking very generally these abilities may be grouped as relating to language or to motor facility, and this broad differentiation is substantiated by comparisons of the ranks of the children in the various class subjects. The average correlations between the different academic subjects are from 0.60 to 0.80, whereas a comparison of academic subjects with hand work gives average coefficients varying from 0.12 to 0.42. However, we can not compare, for instance, the correlation between the Psychomotor Test and woodwork, with the correlation between academic subjects and woodwork and thereby say that success in English is the better indication of the child's future success at hand work than the Psychomotor Test. For there is a high probability of the same influencing factors being present in both the ratings of the academic teacher and those of the woodworking teacher, which cause them to be similar and thus raise the coefficient above that obtained by comparing a test with either set of ratings.

With so little correspondence between proficiency in the academic subjects and proficiency in motor tests or motor work, the academic record alone would be an entirely inadequate basis for guidance to such courses of training. Considering the tests as a basis, we find that it will be necessary to work out combinations of tests diagnostic of capability in each subject. For, especially in the manual training group, the several subjects can hardly be designated as representing exclusively one type of aptitude or the other, nor can one group of tests be applied as indicative of possible success in all lines of manual work. Certain subjects constituting a manual training course demand abilities quite the reverse of each other. The tests correlating with marks in cooking are of general intelligence and language comprehension, while those correlating with sewing are tests of speed and precision of finger movement. Yet at present, a girl choosing the manual training course, automatically takes both of these subjects. A combination of Pintner's Non-language Test and the Psychomotor would form part of a group for prognosing success at woodworking. Bookbinding might require a group of tests similar to those for sewing. The Thorndike test exceeds all others, even those of general intelligence, for the prognosis of general academic success. English is especially related to this test. History, it will be noted, shows excellent correspondence with general intelligence—being particularly high even with the non-language tests.

The above correlations on academic subjects augment similar correlations already made by other investigators. Those on the manual work, I believe, constitute the first group contributing to the solution of the problems of the selection of children for manual training courses. They are offered as a basis for further study. Analysis by means of correlations of the abilities demanded by school subjects will in all probability aid in the better formation of courses of study and in the solution of problems of individual adjustment.

THE EFFECT OF THE STUDY OF LATIN ON ABILITY TO DEFINE WORDS

A. R. GILLILAND

Lafayette College

The value of the study of latin—granting that such value exists—should be as amenable to measurement as are other educational products. More or less successful attempts at such measurements have been reported by Swift, Starch, Partridge, Harris and Foster.[1] The surprising generalization that can be made from these studies is the small amount of increased ability that is to be found in students with Latin training either to get on in other languages or to define words. After necessary deduction has been made for the fact that the organization of our school systems and the prejudice of educated parents have tended to influence the better students to elect Latin, little real difference seems to remain between the Latin and non-Latin groups.

The present study is similar to some of those already referred to with the exception that two attendant factors have been measured where formerly none or, at most, only one has been controlled or measured. The measurement of ability to define words with groups of students that have not studied Latin, and groups that have studied the language for 2, 3, 4, or 5 years was obtained as well as the intelligence scores for these same groups. Also, the general standing in college for the first semester of the freshman year was secured.

One hundred fifteen college freshmen, selected at random, were tested on their ability to define a list of 40 words. The first 10 words of the list were English words of Anglo-Saxon origin; the next 20 words were English words of Latin origin, and the last 10 words were English words of Greek origin. Technical definitions were not required. The list of words is herewith given.

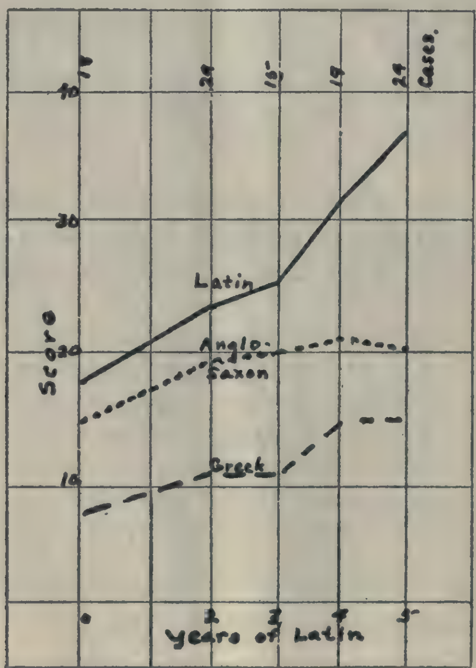
Anglo-Saxon Origin		Greek Origin	
tithe	bier	synchronize	anthropomorphous
midwife	dowery	phenomenon	bibliophile
cog	budget	lithograph	polychrome
broach	bolster	photometer	genesis
cooper	squib	heterodox	pantheism

[1] See brief review of these reports by Daniel Starch in Educational Psychology; 1920, pp. 230–236.

Latin Origin

ossify	impeccable	congenital
omnscience	translucent	sanguinary
hibernate	litigation	extirpate
predatory	moratorium	mendacity
impecunious	quadruped	parricide
adhesion	supernatant	irrefragable
malleable	longevity	

The words were given orally and in order by one or the other of two experimenters. In grading, each correct definition was given three points credit. When the general idea of the word was given but lacking in some more or less essential detail, a credit of two points



Graph I.—Showing the scores on the test in defining words of Anglo-Saxon, Latin, and Greek origin.

was given. Where there was a very hazy idea or mere ability to use a word in a sentence, without a clear notion of its meaning, one point credit was given. On this basis the Angle-Saxon group of words could give a maximum credit of 30 points, the Latin words 60 points, and

the Greek words 30 points, or 120 points credit for correct definitions of the whole list of 40 words. To insure uniformity in grading, one experimenter graded all the papers and the other experimenter made a re-checking to prevent errors or unfair grading.

The accompanying table gives the scores for each subject grouped on the basis of the years the subject had studied Latin. High school and college Latin were combined and since the experiment was conducted in the spring, college Latin was counted as a full year. The scores for each of the three groups of words have been kept separate in the table.

Graph I shows the increase in ability to define English words of Anglo-Saxon, Latin and Greek origin dependent upon the number of years spent in the study of Latin. Not only were those who had studied more Latin able to define more words of Latin origin but they were also able to define more words of Anglo-Saxon and Greek origin appreciably better. One or the other or probably both of two factors account for this. Either the men who had studied Latin for a longer time had a greater native capacity for language, or at least for Latin, or the study of Latin developed a method of attack which helped in defining any type of word. Probably those who studied Latin for several years *did* have more ability for Latin or a greater interest in the subject that the rest, otherwise it would have been dropped. In a large number of cases the study of Latin seemed to have developed a method of attacking new words, illustrated by numerous attempts to break these words up into their component parts. Admittedly the possible greater native ability for Latin in the groups with more years spent on Latin is a complicating factor which can not be isolated nor accurately measured in the present study.

TABLE 1

Anglo-Saxon	Latin	Greek
No Latin		
22	20	13
13	23	7
12	26	8
10	20	10
11	17	5
12	17	9
15	29	17
17	14	3
15	20	1
13	9	2
16	16	1
12	13	0
9	8	3
17	14	5
12	20	5
9	12	5
12	14	7
9	4	5
15	13	9
12	19	16
20	11	10
26	35	13
27	24	15
Average		
14.6	17.3	7.3
Average Deviation		
3.9	5.4	4.0
2 Years Latin		
24	22	16
16	27	12
18	25	7
20	45	20
14	10	6
18	24	9
15	22	8
16	24	7
17	24	12
24	22	10
16	28	15
21	22	12
21	14	5
17	22	9
14	24	8
20	41	11
20	29	12
21	6	7
18	18	14
29	25	17
25	25	10
20	22	8
27	26	8
20	24	13
5	17	6
21	33	14
17	12	3
11	13	6
16	42	18
23	16	9
13	17	5
27	34	29

Anglo-Saxon	Latin	Greek
Average		
18.9	23.1	10.3
Average Deviation		
3.8	6.4	4.0
3 Years Latin		
17	19	5
17	23	8
18	31	7
13	21	12
20	27	8
20	25	5
22	24	9
21	19	7
16	30	16
20	24	12
21	26	14
28	34	17
22	16	11
19	26	11
Average		
19.6	24.6	10.2
Average Deviation		
2.4	3.8	3.2
4 Years Latin		
26	31	13
21	37	22
21	40	12
20	36	10
21	40	13
16	44	17
18	36	12
20	44	28
21	43	18
18	42	15
21	36	25
27	28	9
25	20	12
22	21	9
28	44	12
22	26	12
26	34	20
23	35	14
Average		
20.6	31.5	14.2
Average Deviation		
2.5	8.2	3.8

TABLE 1.—*Continued*

Anglo-Saxon	Latin	Greek
5 Years Latin		
26	47	18
24	52	18
24	49	22
14	19	5
25	46	22
16	42	20
20	36	15
25	44	20
22	44	18
26	42	18
8	20	3
16	30	4
22	40	19
23	38	10
24	47	17
25	33	10
24	51	19
19	34	15
21	31	14
22	32	14
12	29	20
19	26	12
18	25	7
24	42	16
16	18	6
Average		
20.0	36.7	14.6
Average Deviation		
4.0	8.0	4.7
6 Years Latin		
22	51	18
20	39	7
Average		
21.0	45.0	12.5
Average Deviation		
1.0	6.0	5.5

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY CECILE COLLOTON

Department of Educational Psychology, The Lincoln School of Teachers College

Intelligence Tests

The Present Status of Mental Testing. Stephen S. Colvin. Educational Review, 1922, October, 196–206. A summary of the development of intelligence tests and a discussion of the general principles underlying their use.

The Attainment of Pupils on Certain Group Intelligence Tests. Dora Keen Mohlman. School and Society, 1922, Sept. 25, 359–363. Six tables report median scores, Grades VII–XII, for the following group tests; National, Otis, Terman, Indiana, Haggerty and Army Alpha. Three tables report median scores, Grades I–VI, for the Holley Picture Completion, Haggerty Delta 1, and the Illinois Examination. Data are taken from 22 investigations.

Undeveloped Resources: Some Studies in Group Intelligence in Sioux City High School. Robin Lynn Hamilton. School and Society, 1922, Oct. 7, 416–420. Eleven studies of the mentality of high school students as shown by the Army Alpha test and its relation to achievement, failure, elimination, etc.

Attempts at Test Validation. Raymond Franzen. Journal of Educational Research, 1922, September, 145–158. An investigation of 14 intelligence tests with reference to their correlation with certain defined criteria. Pleads for fewer tests and better construction.

Can Teachers Select Bright and Dull Pupils? G. F. Verner. Journal of Educational Research, 1922, September, 126–132. A comparison of the upper and lower per cent of 286 children in the schools of St. Paul, Minnesota as selected by the teachers with the intelligence quotients as determined by group intelligence tests. Teachers estimates unreliable.

The Mental Age of Adults. Edward A. Lincoln. Journal of Educational Research, 1922, September, 133–144. Is the mental age of adults 16 years as assumed by Terman, or 13½ as shown by the army results? Discusses Terman's objections to the inferences from the army testing.

Some Reactions to Standardized Tests. Emilie V. Jacobs. The Journal of Educational Method, 1922, September, 33–36. A grouping of VII*B* and VIII*A* pupil on the basis of Haggerty Intelligence Test IQ's. Opinions of eight teachers on the value of such a classification.

A Mass Mental Test for Use With Kindergarten and First Grade Children. Clara H. Town. Journal of Applied Psychology, 1922, June, 89–112. Description of a non-verbal test for use with young children. The "Picture Game" is reproduced in full and complete directions for its use are given as well as experimental results.

The Selection of a Successful Secretary. A. T. Poffenberger. Journal of Applied Psychology, 1922, June, 156–160. Results of the trail of the Army Alpha intelligence test in a secretarial school.

A Glimpse of High School Courses as Measured by the Otis Test. Ruth S. Clark. Journal of Applied Psychology, 1922, June, 185–191. A study of the Academic, the Commercial, the Technical, the Industrial Arts, and the Dressmaking Courses in High School shows various levels of intelligence within the groups.

Occupational-intelligence Standards. Douglas Fryer. School and Society, 1922, Sept 2, 273–277. A classification of 96 occupational designations on the basis of the average intelligence score as determined by Army Alpha and "Business Alpha." Of value for vocational guidance.

Educational Tests

Scores Made by Seniors on the Hotz Algebra Scales Compared with Scores Made by High School Students Taking Algebra. Clifford Woody. School and Society, 1922, Sept. 9, 303–306. Seniors retain a relatively large amount of knowledge of the formal aspect of algebra but fall very low in the ability to solve written problems.

Recent Developments in Silent Reading Tests. C. R. Stone. Journal of Educational Research, 1922, September, 102–115. Part I—a discussion of three types of silent-reading tests with illustrations. Part II—a detailed account of the construction and use of the Stone Series of Narrative Reading Tests.

Miscellaneous

Subsequent History of E—; Five Years After the Initial Report. L. S. Hollingworth, C. G. Garrison and Agnes Burke. Journal of Applied Psychology, 1922, June, 205–210. Mental and physical measurements and scholastic achievements of a child with an IQ of 187.

Repetition versus Other Factors in Learning. J. W. Barton. The Pedagogical Seminary, 1922, September, 283–287. Urges more attention to (1) readiness (native and acquired nature), (2) stimulation in keeping with this nature, and (3) knowledge of results through objective checks.

The Social Purpose of the Education of the Gifted Child. George S. Counts. Educational Review, 1922, October, 233–244. Importance of the development of a strong sense of social obligation in the gifted child. General discussion of the present day trend in the education of children of talent.

Teachers vs. Mental Tests as Prophets of School Progress. Garry C. Myers. School and Society, 1922, Sept. 9, 300–303. Argues that correlation with school progress is not a true measure of the validity of an intelligence test. The average teacher prophesies future progress of pupils better than intelligence tests.

Improving the Reading Ability of College Students. Cliff W. Stone. The Journal of Educational Method, 1922, September, 8–23. An investigation of the reading ability of fifteen college classes. Methods of improvement used are described and results discussed.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

1. *Physical Training Psychologized.*—Two volumes[1] covering two distinctly separate aspects of physical training are included under this head. The first is concerned with formal exercise and corrective gymnastics. The author sets forth in forceful terms the psychological premises underlying physical training. He says, "In the past physical training has been largely esoteric. Its meaning, except in a most general sense, was hidden as if it were beyond the comprehension of the ordinary mind." He has endeavored to give physical training teachers suggestions which are exceedingly simple, manifestly productive of result and honestly, completely and powerfully true. He conceives the purpose of physical training to be the maintenance of good health, increased vigor, mental and physical efficiency and the promotion of neuro-muscular and psycho-motor education. In a historical statement he relates how dissatisfaction with prevailing practice led him in 1902 to determine: First, the definite results or objectives of physical training, to classify these and determine their relative worth; next to ascertain by what exercises and methods of instruction these ends could best be met; then to devise a comprehensive plan and test it by practice under varied conditions. During 20 years the suggestions have been put into practice by an ever widening circle of instructors. After reading this book, physical training teachers should have a keener realization of the principle underlying the success of their instructional activities and a sound technique for securing the desired pupil reactions. Those who are more specifically concerned with young children are referred to the other book, a volume contributed by an experienced playground worker. Observations and interpretations of the spontaneous and supervised play of children are assembled in an attempt to clarify the underlying philosophy and the psychological implications.

[1] Ward, Crampton C., M. D.: "The Pedagogy of Physical Training." The Macmillan Co., New York, 1922, pp. XV + 257 and Sies, Alice Corbin: Spontaneous and Supervised Play in Children. The Macmillan Co., New York, 1922, pp. XII + 442.

This book is admirably suited for use in connection with courses for prospective teachers and playground leaders, because of the abundance of illustrative material, the selected list of collateral readings, with topical references, textbook assignments, questions, and exercises included in the appendix. L. Z.

---

2. *An Experimental Study of the Reading of Numerals.*[1]—To the already impressive list of reading monographs from the University of Chicago laboratories must be added Terry's study of the eye movements involved in reading isolated numerals and numerals incorporated in sentences or problems.

Part I reports four preliminary studies by introspective methods. A number of graduate students acted as subjects. For Part II photographic apparatus was used to record the eye movements of six adult subjects, three of whom had participated in the preliminary studies. Each subject read simple arithmetical problems, isolated numerals of various lengths and ordinary prose. The educational implications of the data are such that the present study is to be considered as introductory to more elaborate detailed inquiry into the development of the specific eye habits involved.

Of the conclusions drawn the following are particularly significant: Ordinary prose is read much faster than either the problem material or isolated numerals. Problems are invariably re-read in whole or in part before computation ensues. In the first reading only short, common or round numbers are read. The first reading is usually for the purpose of ascertaining the conditions of the problem. Partial first reading of the numerals is conducive to a quicker grasp of the meaning and conditions of the problem. The final chapter is a valuable discussion of the numerous practical applications of the conclusions to classroom teaching and contains definite recommendations which should find their way into practice. L. Z.

---

3. *A Significant Contribution to the Psychology of Reading and Spelling.*[2]—This volume reports an exceedingly intensive diagnostic

[1] Terry, Paul Washington: How Numerals are Read. *Supplementary Educational Monograph*, No. 18. Department of Education of the University of Chicago, 1922, pp. XIII + 109.

[2] Gates, Arthur I.: The Psychology of Reading and Spelling with Special Reference to Disability. *Contributions to Education*, No. 129, Teachers College, Columbia University, 1922, pp. VII + 108.

investigation with far-reaching implications set forth in an unusually thoroughgoing interpretation of results. Over a hundred children in the classes of a single school were put through an inclusive battery of group tests and individual examinations covering mental ability, achievement in school subjects, reaction to visual stimuli and spoken words, and sensory and motor reactions. The results were subjected to the most careful statistical analysis. The technique used to discover causes of inability furnishes an approach to some of the vital problems which must be solved before we dare hope for scientifically determined remedial work. The serious reader cannot fail to be convinced of the urgency of a psychological analysis of school functions, and gets a new vision of the nature and scope of diagnostic research.

The critical discussion of the mooted subject of phonics is constructive, and certainly throws new light on the subject. Disagreement with the conclusions of other workers concerning the significance of eye movement habits calls for the reconsideration of the evidence and more careful discrimination between causes and effects.

Distinct relationships between perceptual phases of reading and spelling are shown, and numerous other factors contributing to spelling disability are discussed. In this connection the study shows that, in the nature of the case, the numerical statement of correlation may conceal rather than reveal important facts, because the causal relations may be different in the extremes of the distribution, while the correlational methods assume rectilinear regression.

The relative infrequency of actual cases of serious disability due to defects of the central nervous system is noted, and other causal factors are discussed under the following headings: (*A*) Unfavorable training and environmental influences, (*B*) Unfavorable behavior of a general character, (*C*) Defects of the sensory mechanisms, (*D*) Defects of motor mechanisms, (*E*) Defects of connecting mechanisms.

In conclusion Doctor Gates says:

> A case of inability to read affords frequently a tangle of difficulties that experts from several professional fields working together may be unable to disentangle. Such a situation portrays clearly the need of a new group of specialists who will make the solution of such problems their main work. It will demand a mastery of the knowledge and technique of several sciences. Such research is essential, not only because it is plainly desirable to diagnose and remedy the conditions underlying disability, but because the development of general methods of instruction depends upon such knowledge as these achievements will provide.

There are those who expect the average child to blossom into reading. The final sentence in this significant study applies here—

> It is folly to expect children to learn functions as complex as reading and spelling economically and effectively without instruction, and is equally futile to attempt to devise adequate methods of instruction without intimate knowledge of the constituents of these functions and the influence of a variety of factors upon them.

---

4. *A Critical Study of Certain Silent Reading Tests.*[1]—This bulletin has to do with the determination of the comparative validity and reliability of some of the standardized silent reading tests which purport to measure rate and comprehension. One cannot help but wonder why in 1922 the Ayres-Burgess, and the Haggerty tests were not included and why certain other tests were included. These vagaries of selection make the conclusions inconclusive, regardless of the adequacy of investigational technique. The author in his preface suggests that the monograph may be of interest to students in the field of educational measurement. In this connection the bulletin may possibly be of help to instructors who are looking for type-studies to illustrate such problems as pertain to reliability, validity, constant and variable errors, and the correction of errors due to sampling.

L. Z.

---

5. "*After Tests, What Next?*" This is the question which the schoolman in 1922 is asking. He has now passed by the milestones of "What are intelligence tests?" and "Do the tests really measure intelligence?" It is with hope of suggesting to the educator what to do after he has given intelligence tests, that this book[2] has been published. It is a composite by many authors, showing different kinds of school reorganization prompted by the results of intelligence testing. So far most of these suggestions have been scattered here and there in various journals, and it is, therefore, useful to have the results of some definite "next steps after testing" bound together in one volume.

---

[1] Monroe, Walter S.: A Critical Study of Certain Silent Reading Tests. *Bull.* 8, Vol. XIX, No. 22, Bureau of Educational Research, College of University of Illinois, Urbana, Ill., pp. 52.

[2] Terman, L. M.; Dickson, V. E.; Sutherland, A. H.; Franzen, R. H. and Fernald, G.: Intelligence Tests and School Reorganization. *Subcommittee Report, N. E. A.* World Book Co., 1922, pp. VIII + 111.

In no sense, however, does the present volume attempt to cover or, indeed, indicate all the possibilities which a testing program may open up. Neither is it a compendium of all the experimental work done up to the present time. Nevertheless, it is very suggestive and helpful.

In a capital introductory chapter Terman, as editor, sets forth some of the results of the testing movement up to the present time. He reminds us of the recency of the intelligence test. He indicates its rapid growth, giving an estimate of over two million children tested in 1920–21. He shows briefly the heterogeneity of mental ability of pupils in the same grade at the present time and argues for a policy of homogeneity of intelligence in grades by means of a multiple track system wherever possible.

The other contributions deal with specific work in various parts of the country. The several types of classes developed at Oakland are described by Dickson. He shows what is being done for the bright and for the dull child. With reference to the latter group he indicates how they are being moved on into the junior and senior high school by means of a reorganized curriculum to suit their special needs. To make the mentally slow child repeat again and again the work of a given grade is certainly not a solution of his difficulty. The brighter children are being handled so far as possible in special classes, and both plans of enriching the curriculum and increasing the speed are being tried.

Sutherland describes the Adjustment Rooms of Los Angeles and emphasizes the necessity of adjusting the curriculum to the mental level of the child. Franzen describes briefly the Accomplishment Quotient and gives some data from his Garden City experiment. He rightly emphasizes the need for making a better use of the intelligence of the brighter children in our schools. Tupper gives an account of the use of intelligence and educational tests in a small city school system, while Fernald gives a very brief account of her interesting work with children who have difficulty in spelling and reading. All the six chapters of the book are valuable and stimulating. There is, however, no logical connection between the various chapters. The book is merely a collection of articles by different writers. Nevertheless, the topics are timely and are well worth the attention of all teachers and educators.

R. P.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Vol. XIII December, 1922 No. 9

WHAT SHALL WE EXPECT OF THE AQ?

HERBERT A. TOOPS

AND

P. M. SYMONDS

Institute of Educational Research

Teachers College

The AQ, or accomplishment quotient, procedure is one of the most recent, not to say most promising, acquisitions of the educational psychologist. Its implications have been subject to no little confusion even among the originators of the technique, as will be evident when we consider its derivation in more detail below.

The educational problem of motivation, as a partial solution from which the AQ hypothesis was derived, involves in its widest aspects the problems of the ultimate aims of education; the administrative problems of sectioning, retardation, promotion, and elimination; the pedagogical problems of motivation and differential treatment; and the research problems involved in the measurement of educative capacity and of educational product.

We may or may not agree to delegate to the educational philosopher the task of determining the ultimate aims of education. By unanimous agreement, the realization of those aims is at present left to the school administrator, the pedagog and the educational psychologist. By pointing out some of the limitations, as well as advantages, of certain postulates and "axioms" connected with the AQ hypothesis, the authors hope to show the probable effect of the hypothesis upon administrational and testing practices, to point out some new ends or aims of education that may soon come to the foreground of public discussion, and possible resulting changes in school administrational procedures. Thus the authors will raise many questions without attempting to answer them adequately, if at all. The man of practical affairs may be inclined to remonstrate that our remarks are

destructive criticism rather than suggestive of constructive programs. On the contrary, the authors hold to the point of view that an awareness of our ignorance is of as much value in pointing out the road to progress as a knowledge of our accepted scientific "truths."

A Comparison of Three Current AQ Concepts

The accomplishment quotient is advocated by Franzen,[1] and under the name "achievement quotient" by Monroe and Buckingham,[2] as a measuring device for combining in an effective way the results of educational and mental tests into a measure of *educational achievement relative to the pupil's capacity to progress.* Essentially the same purpose is served by a different statistical technique developed by Pintner.[3]

The AQ is to be considered as the "degree to which a pupil's actual progress has attained to his potential progress by the best possible measures of both,"[4] or as a "simple method of comparing a pupil's achievement age with his mental age (learning capacity)."[5] Apparently there is no single word in the English language which adequately expresses what is meant by the term AQ. In its statistical derivation *it is quite as abstract a concept as π or r.* Its formula is:

$$\mathrm{AQ} = \frac{\mathrm{EQ}}{\mathrm{IQ}} = \frac{\frac{\mathrm{EA}}{\mathrm{CA}}}{\frac{\mathrm{MA}}{\mathrm{CA}}} = \frac{\mathrm{EA}}{\mathrm{MA}} \tag{1}$$

where,

AQ = Accomplishment, or achievement, quotient
EQ = Educational quotient
IQ = Intelligence quotient
EA = Educational age
CA = Chronological age
MA = Mental age.

[1] Franzen, R.: The Accomplishment Quotient. *Teachers College Record,* Vol. 21, No. 5, Nov., 1920, pp. 432–440.

[2] Monroe, W. S. and Buckingham, B. R.: Illinois Examination. *Teachers Handbook,* University of Illinois, Bureau of Educational Research, July, 1920, p. 31.

[3] Pintner, R. and Marshall, H. A.: A Combined Mental-educational Survey. *Journal of Educational Psychology,* Vol. 12, No. 1, Jan., 1921, pp. 32–43.

[4] Franzen, R.: *loc. cit.,* p. 436.

[5] Monroe, N. S. and Buckingham, B. R.: *loc. cit.,* p. 11.

Pintner's method consists essentially of transmuting educational test and mental test scores into index values ranging from 0 to 100 for a given age, average ability in each being 50. The method assumes a normal distribution of both mental and educational talent. Pintner uses as his measure of motivation:

$$\text{Difference} = \text{Educational index} - \text{Mental index}. \qquad (2)$$

This measure is "the difference between a pupil's native capacity and his actual accomplishment."[1]

According to Franzen, an AQ of 1.00 indicates "optimum accomplishment" or "what a pupil is able to do under the best conditions;" and according to Monroe and Buckingham, it means "that the pupil has achieved exactly as well as the average of the pupils of his mental age;"[2] while, according to Pintner, an index difference of zero, occurring when the mental index is equal to the educational index, or a corresponding AQ of 1.00, apparently means that the pupil is doing educationally exactly what "is usually accomplished by children of like mentality."[3]

According to Franzen, an AQ less than 1.00 means that the pupil is doing school work which is less than normal for his mentality, and, according to Monroe and Buckingham "if a pupil's achievement quotient is 0.75, we have evidence that he has achieved only 75 per cent as much as the average of the pupils of his mental age;"[2] while, according to Pintner, "a minus difference means that the child is doing less educational work than he has the ability to accomplish"[3] although, as noted below, he does not imply that a plus difference indicates that the pupil is doing more work than he has ability to accomplish.

According to Franzen an AQ of more than 1.00 is impossible, as represented in his statement: "One's differences when EQ is subtracted from IQ are always positive when they are large enough to be significant and small enough to seem spurious when they are negative . . . It is safe, therefore, for practical use to assume that the optimum accomplishment is 1.00;"[4] and according to Monroe and Buckingham, "If the pupil's achievement quotient is 130, it means that he has achieved 30 per cent more than the average of the pupils

[1] Pintner, R. and Marshall, H.: *loc. cit.*, p. 37.

[2] Monroe, W. S. and Buckingham, B. R.: *loc. cit.*, p. 11.

[3] Pintner, R. and Marshall, H.: *loc. cit.*, p. 38.

[4] Franzen, R.: *loc. cit.*, p. 436.

of his mental age,"[1] while according to Pintner, "a plus difference means that the pupil is doing more educationally than has usually been accomplished by children of like mentality."[2]

Pintner, and Monroe and Buckingham find many pupils making "more than average accomplishment for their mental age," Pintner specifically stating that "it is useless to attempt to set up any such ideal standard (of what ought to be accomplished under ideal conditions where each child is working up to the limit of his capacity); in contradistinction, Franzen states, "we can measure the approximation to ideal educational performance of any one child in any one elementary school subject through the approximation of this accomplishment quotient to 1.00." It is evident that even among the originators, there is a great difference of opinion in regard to the meaning to be attached to any AQ. Part of this disagreement, no doubt, will be eliminated once all compute their indices in identical statistical fashion, and on the same tests.[3]

The Disagreement in Terminology Involved in the AQ Hypothesis

Dr. Otis has pointed out to us the errors of terminology in which we are likely soon to be involved in regard to the various ratios. We find research workers talking of a *reading quotient*, that is, of the ratio of reading-subject-matter age to chronological age; of a *reading-accomplishment quotient*, that is, the ratio of reading-subject-matter age to mental age; and finally of a more general accomplishment quotient in the sense of the ratio of average subject-matter ages in a number of school subjects to mental age. By early agreement, research workers may decide upon adequate definitions of standard terms and thereby prevent ultimate hopeless confusion. It may be noted that the term used for the more general accomplishment quotient must be defined in terms of *the subject-matter ages to be included while also taking into account how they are to be combined or weighted* if we are to hope for even approximately valid comparisons of the work of various research workers. A pupil's general AQ evidently depends in a very real way upon his election of subject matter and so cannot be

[1] Monroe, W. S. and Buckingham, R. B.: *Loc. cit.*, p. 38.

[2] Pintner, R. and Marshall, H.: *loc. cit.*, p. 38.

[3] Part of the confusion is due to the fact that Franzen used the Stanford Revision Individual Test while the other investigators used Group Intelligence Tests.

expected to be as constant even as his IQ. Likewise, mental age needs to be defined in terms of what tests shall be used and how they shall be combined if we are to hope for any reasonably comparable results in intelligence measurement from different research workers, or even from the same research worker in successive instances. In this article, unless specifically noted otherwise, we have taken the term "accomplishment quotient" to mean either the subject-matter-accomplishment quotient or the more general accomplishment quotient. Our discussion of the limitations of the Q procedure will hold for either the more specific or the more general case.

Difficulties Involved in the Norms Used, and in the Selection of Standard Tests

One cause for confusion in the interpretation of the meaning of a given AQ lies in the difference in procedure used in computing norms. Monroe and Buckingham, and Pintner, follow the customary procedure in determining a norm; namely, finding that score which is the median for a given age and calling that score the norm for the age. Franzen finds the average age of all people who make a given score, thereafter calling the given score the norm for the average age thus found. Thus, stated in statistical terminology, the former workers make use of the regression of score on age, while the latter makes use of the regression of age on score. The regression of age on score, it will be noted, is the customary regression line used in such problems as that of predicting the age at death from a statistical measure of the person made prior to the event. The adoption by all workers of the other regression line, if proven statistically advisable, is not an impossible task. We point out below that the use of the other regression line in norms does not do away with what seems to us to be a very real objection to the AQ procedure.[1]

The question of equivalence of scores on tests constructed by different research workers is also in a state of flux, as are many of the statistical implications of mental and educational measurements of which the controversy regarding the two regression lines in norms is but one example. Otis is advocating the use of a line, which when plotted lies between the two regression lines for converting mental test scores of one scale into "equivalent scores" on the other, disregarding the

[1] Recent reports show that local community selection is so great that "blanket" norms are often meaningless. See Chapman below.

fact that there is no true equivalence of two test scores.[1] Without true equivalence of different mental and educational scales, we cannot expect identity of interpretation of AQ's secured by different workers using different mental test, educational tests, or both. We are continually being reminded nowadays that the IQ was devised as a brightness measure for one intelligence scale, the Stanford Revision of the Binet Scale; and that, consequently, the IQ procedure is not in strict scientific usage applicable to other scales than the Stanford scale.[2] If the AQ procedure is to have a monopoly on Stanford IQ's, it necessarily must have a monopoly on Stanford MA's, for it will be seen that the CA's cancel out in equation (1), leaving only two simple variables, EA and MA. We need but one of these invalidated in order to have the whole fractional equation invalidated.

And whose EQ shall be considered a standard one? Not only does this point to an inadequacy of the AQ procedure but of the IQ and EQ procedures as well. There is good reason for believing that the IQ is not the best possible brightness measure, even in the case of the Stanford Scale. As hinted at by Toops and Pintner[3] there are an infinite number of comparatively simple equations of the first degree—not to mention higher degrees—of the type,

$$\mathrm{IQ} = \frac{(\mathrm{MA}) + k}{(\mathrm{CA}) + k} \qquad .(3)$$

which will fulfill the requirements of yielding: (1) A ratio of 1.00 for perfectly normal individuals, (2) ratios of more than 1.00 for individuals brighter than normal, and (3) ratios of less than 1.00 for individuals who are duller than normal. A particular one of this family of curves may fulfill better the additional desirable requirement of approximate constancy through the grade-schools ages than does the present IQ formula. The IQ equation can be thought of as the simplest possible case of the more generalized mathematical ratio,

$$\mathrm{IQ} = \frac{(aM^n + bM^{n-1} + cM^{n-2} + \ . \ . \ . + M) + K}{(a.C^n + b.C^{n-1} + c.C^{n-2} + \ . \ . \ . + C) + k} \qquad (4)$$

where M equals mental age, and C equals chronological age. Among

[1] Thorndike, E. L.: On Finding Equivalent Scores in Tests of Intelligence. *Jour. of Appl. Psych.*, Vol. 6, No. 1, 1922, pp. 29–33.

[2] Trabue, M. R.: Some Pitfalls in the Administrative Use of Intelligence Tests. *Jour. of Educ. Research*, Vol. 6, No. 1, 1922, pp. 1–11.

[3] Toops, H. A. and Pintner, R.: Curves of Growth of Intelligence. *Jour. of Exp. Psych.*, Vol. 3, No. 3, 1920, p. 235.

the "infinite-infinite" number of possible equations, there is probably one which will fit the empirical facts better than the one now used.

Technical Difficulties in Securing Alternative Standards of Capacity and Attainment

Fairly comparable IQ's, EQ's and AQ's *can only be obtained by at least taking into account the reliability coefficients of the* different educational and mental tests respectively. That is, concretely, an IQ secured by the Jones Mental Test can only hope to measure exactly the same thing as an IQ secured by the Johnson Mental Test if the Jones Test correlates perfectly with the Johnson Test. Not even equal correlation with the same identical criterion of intelligence solves the problem. As an illustration, *suppose two tests*, 1 *and* 2, *are totally uncorrelated with each other, they will yet correlate each with a valid criterion of intelligence to the maximum extent of* 0.71. This may be shown by substituting the values, $r_{12} = r_{13}$, and $r_{23} = 0$, in the formula for the multiple correlation coefficient involving three variables when the multiple correlation coefficient is a maximum, or 1.00. Thus:

$$1.00 = r_{1\cdot 23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}} \tag{5}$$

Substituting the above values,

$$1.00 = 2r_{12}^2, \text{ whence } r_{12} = r_{13} = \sqrt{.50} = 0.71$$

If it were possible to construct two "intelligence" tests, a group test and an individual test, which would correlate zero with each other, both might yet correlate equally with a "valid criterion of scholarship" as highly as 0.71; and yet on the one test an idiot would as likely as not be rated genius, and *vice versa* which argues neither for the group test nor for the individual test. This problem must be settled on another basis than statistical theorizing since it is practically impossible to design two tests according to the above specifications. To determine socially *valid* and *comparable* AQ's we must consider validity, correlation with an adequate criterion, as well as correlation between the two intelligence and the two educational tests used by two different research workers. The two forms of test may be perfectly reliable and yet not measure at all what we would have them measure. The conclusion is obvious. We need not a commercialized multiplication of scales and an equally thoughtless diversity of

statistical techniques but an ultimate soundness of method. The very equation of the AQ, if written in another form

$$\text{EA} = (\text{AQ})\ (\text{MA})$$

$$\begin{pmatrix}\text{educational}\\ \text{achievement}\end{pmatrix} = \begin{pmatrix}\text{educational}\\ \text{environment}\end{pmatrix}\begin{pmatrix}\text{mental}\\ \text{capacity}\end{pmatrix}$$

means that educational achievement is equal to mental capacity as it is acted upon by an educational environment (motivator) which varies in its intensity for individual pupils from somewhat more than zero AQ to somewhat more than 1.00 AQ. It would be but mockery to say that any one of our multitudinous intelligence scales measures "mental capacity" when scarcely any of them correlate highly with each other, none correlates highly with the social acts wherein intelligence functions, and some do not correlate very highly even with themselves in their alternative forms. A ratio of such unreliable variables is necessarily less reliable than either of its components.

In securing his measure of capacity, Pintner uses non-verbal intelligence tests "to get as far away from language and the things taught in school" as possible. Yet his educational norms are still based on what is now taught. Pintner quite rightly wishes to get an "ultimate measure of ability or rate of doing work" which he hoped to get in non-language tests. It is known, from the work of Herring,[1] Gates, the N. I. T. tests, and others, that non-language intelligence tests do not correlate nearly so well with ability to get along in school (as the school subjects are now taught) as do verbal tests. It has been found that the more verbal the tests the higher their correlation with an "adequate" criterion of intelligence or of ability to get along in school. We need only consider the limiting case of an intelligence test which is so "non-verbal" as to correlate zero with achievement in order to see that the *measure of capacity must correlate highly with the measure of attainment.* The real requirement is that the test used shall be as little susceptible as possible to improvement through practice or coaching.

The Experimental Group Which Determines "Capacity" Norms Should be Maximally Motivated

Without testing "maximally motivated" pupils to determine our norms of "potentiality," we can but approximate an ultimate measure

[1] Herring, J. P.: Verbal and Abstract Elements in Intelligence Examinations. *Jour. of Educ. Psych.*, Vol. 12, No. 9, 1921, pp. 511–517.

of the capacity to do school work. We need not deny our slow but steady progress in measurement.[1] It will do no harm, however, to realize that so long as our mental tests correlate with an "adequate criterion" to the extent of less than 0.71, "all other unrelated factors" will correlate with the same criterion to a greater extent than 0.71; and further, that "a composite of all other unrelated factors including also some factors common to the first mentioned test," such a composite as it would be impossible to approximate in an almost "totally different" type of test, would correlate with this same criterion considerably in excess of 0.71. Not until we construct intelligence tests which will correlate to the extent of 0.87 with such a criterion will we reduce to half the standard deviation of the criterion its standard error of estimate.

Evidently there is no method whereby *statistically* we can determine when a child is maximally motivated. The best we can do is to arrange an experimental class with the best conditions and incentives to maximal effort that the best pedagogical judgment can devise and then measure what educational product is produced. The person who can arrange greater incentives in a subsequent experiment will secure a greater educational product. "Maximal motivation without neglect of essential school activities" would yield the best norms—"balanced" norms. Scientific method requires that such a group be used as the experimental group in constructing the tests of capacity and achievement. Even then we are in exactly the same position as the time study men of industry who decide that a fair day's work is what the average man produces. In the long run what is considered fair is what the workers will agree to accept as a fair day's work; it is

[1] Pressey, S. L.: Suggestions Looking toward a Fundamental Revision of Current Statistical Procedure as Applied to Tests. *Psych. Rev.*, Vol. 27, 1920, pp. 466–472.

Ruml, B.: Reconstruction in Mental Tests. *Jour. of Phil., Psy. and Sci. Meth.*, Vol. 18, No. 7, 1921, pp. 181–185. (A criticism of Pressey above.)

Pressey, S. L.: Empiricism versus Formalism in Work with Mental Tests. *Jour. of Phil., Psy. and Sci. Meth.*, Vol. 15, No. 16, 1921, pp. 393–398. (The reply to Ruml's criticisms.)

Ruml, B.: The Need for the Examination of Certain Hypotheses in Mental Tests. *Jour. of Phil., Psy. and Sci. Meth.*, Vol. 17, No. 3, 1920, pp. 57–61.

Kelley, T. L. and Terman, L.: Dr. Ruml's Criticism of Mental Test Methods. *Jour. of Phil., Psy. and Sci. Meth.*, Vol. 18, No. 17, pp. 459–465. (Reply to Ruml directly above.)

Chapman, J. C.: Some Elementary Statistical Considerations in Educational Measurements. *Jour. of Educ. Research*, Vol. 4, No. 3, 1921, pp. 212–220.

not necessarily a scientifically determined quantity of work in spite of its scientific appearance, its abundance of fine but unreliable measurement.

Some Curious Phenomena Noted in the Use of the AQ

Let us inquire into the educational treatment accorded subjects of low AQ by Franzen. We see "unmotivated" pupils, discovered by the AQ procedure, given special educational treatment with the general result that their AQ's are brought up to 1.00, but not beyond it. What may be one explanation of this case? It has repeatedly been shown by Kirby, Chapman and others that motivation leads to distinct improvement. In fact, Thorndike[1] tells us that there is no reason to believe that in many functions the acceleration of improvement within the ordinary physiological limits need be a negative one provided *we furnish sufficient motivation.* Franzen used the Stanford Test while the other workers used group tests; he likewise used a different regression line in computing his norms. Aside from these differences, is it not likely that his subjects did improve up to the given expected point, the goal of 1.00 AQ, and that then improvement did stop with few going beyond AQ's of 1.00, because the teacher and pupils were led to believe that an AQ of 1.00 was satisfactory; that is, that the "motive" to improve was greatly lessened or suddenly became of zero value as soon as an AQ of 1.00 was reached? If, as will be shown shortly, half or more of the dull pupils can expend more than "normal effort," why cannot *all of humanity* do more than an AQ of 1.00? It probably can! Why, then, if sufficient incentive is provided, should not at least half of his school system do more than the *average amount of school work usually done by people of the same mental age in school systems in general?* Does not the greatest value of the AQ, after all, consist not in its measuring value but in its incentive value—its value in getting the teacher and pupil *interested in progress?* The graph of progress is a very real incentive to the pupil—so very effective because it compares his educational attainment with himself; because he is competing with himself, and is not required to beat out pupils of greater ability. For, even if of low IQ and he works up to an AQ of 1.00, he is doing "just as well" as the pupil of greater intelligence who does more work and achieves a greater EA. Is mental capacity to be likened to the

[1] Thorndike, E. L.: *Educational Psychology*, Vol. 2, 1913, p. 257.

fabled beggar's wallet which can be filled only so full before it will be filled to overflowing and burst from its very opulence? And, yet, may it not be good school policy for the present to keep the AQ from going above 1.00, in order to insure that the school will not put too much emphasis merely on the things which the tests measure, and allow opportunity for securing some of the appreciations or attitudes which, though intangible, are valid objectives of education? The EQ aims only to measure education as it now is, and not as it ought to be. Granted the truth of the hypothesis, the AQ procedure is a very real incentive method, as shown by the fact that the correlation of about 0.6 between EQ and IQ at the beginning of Franzen's experiments was pushed to about 0.9 by intensive stimulation of his pupils to effort.

Another curious phenomenon, noted by Pintner, is that there are more bright people not working "up to capacity" than dull ones who are "doing more than is expected on the average of pupils of their mental capacity." Another investigator in an unpublished report finds a correlation of—0.40 between MA's and AQ's in the case of pupils of Grades V to VII. Is there not significance in this fact which we may interpret from the known facts of the school situation itself? Is it not a remarkable coincidence that the "below normal" in intelligence are for the more part *above* average in motivation while the "above normal" in intelligence are for the most part *below* average in motivation? We are often inclined to accept the generalization that all good things are positively correlated; correlation and not compensation is the rule in human nature. Either human nature is perverse in its schoolroom duties, or the school methods are badly at fault.

Both statistical methods evidently assume that the *immediate* ideal in education should be to raise the AQ of all "poorly motivated" pupils to 1.00. This makes the statistical assumption that all pupils in all school subjects of a "perfectly adjusted and maximally motivated" school should have an AQ of 1.00; or, all plotted points would lie on a straight line of regression when the subject-matter ages in a given school subject are plotted against mental age; or that, stated differently, in a properly motivated school working up to maximal capacity, the correlation between mental age and subject matter age is 1.00. There is much empirical evidence which will cause us to doubt this ultimately perfect correlation. In the most highly correlated of physical *sizes* of bilateral members of the human body, such as the length of the right arm correlated with the length of the left

arm—not to mention the lesser correlations of the physical capacities of such bilaterally symmetrical members—the correlation is always somewhat less than unity. Should we then expect a perfect correlation between mental capacities? Franzen finds that remedial educational measures recommended for pupils with AQ of below unity brought a majority up to unity, and as above pointed out, he believes that AQ's above unity are spurious, at least when using his methods of computation. The same educational procedure applied to the pupils already at an AQ of 1.00 might have produced some very large AQ's.

A partial explanation of this curious phenomenon will now be presented. It seems likely that the AQ results are *in part* due to the statistical assumptions underlying the accomplishment index technique rather than to an ultimate soundness of method. If we assume, for the moment, that there is not a perfect correlation between educational age and mental age, in a "maximally motivated school" (one in which teaching, school environment and "effort" are ideal) educational index regresses upon mental index and *vice versa*. As will be shown below, pupils of high IQ are then more likely than not to be lower in EA than their MA "would warrant," and conversely, pupils of low IQ are more likely than not to have an EA higher than their MA "would warrant."

If we assume normal correlation of the EQ and IQ, when r is less than 1.00, we obtain a surface of distribution of the two, when plotted against each other, similar to that of Fig. 1. The line *cc*, drawn at an angle of 45° to the horizontal, would represent a line of perfect correlation; *i.e.*, a line on which would be plotted all persons whose EQ's exactly equal their IQ's. Consequently all persons happening to fall on this line have AQ's exactly equal to 1.00. Any person lying above this 45° line, is found to have an EQ greater than IQ, and will therefore have an AQ greater than 1.00. Such a person is P. Conversely, any person lying below this 45° line has an AQ which is less than 1.00. Such a person is Q.

The regression of y on x, or yy, is so drawn that it bisects each vertical array. Its equation is $y = r_{xy}.\frac{\sigma y}{\sigma x}.x$. Let us now consider any given vertical array of people having IQ's less than 1.00; *i.e.*, subnormal in mentality. Such an array will be *any* vertical array to the left of the average (Mx) of the IQ's such as is represented by *hkls*. Half of the area of the array is included above the line of regression of y on x; that is, 50 per cent. in *jklm*. But the area of persons in the array with AQ's greater than 1.00 is represented by the area, $ikln = jklm + jmni = 50$ per cent. of the array + the area, *jmni*. That is: *more than half* (as represented by the excess area, *jmni*) of any unselected dull persons have AQ's greater than 1.00 solely by reason of geometrical necessity, irrespective of whether the hypothesis that EQ can be brought up to IQ is ultimately sound or not; that is, a correlation surface of less than 1.00 always has such an area as *yoc*. Conversely, by consideration

of any vertical array *hkls* of brighter than average persons, more than 50 per cent. have AQ's *less than* 1.00 solely by reason of geometrical necessity, as above. It then follows that (1) the measuring validity of the AQ is thus far an unproved postulate, and that (2) the demonstrated bringing of low AQ's up to 1.00 by special

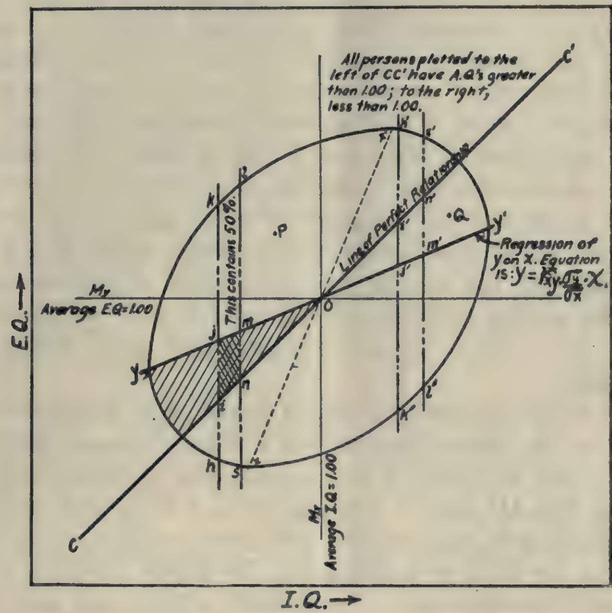


FIG. 1.

educational treatment is no proof of a fundamentally perfect relationship between mental ability and educational achievement. As a corollary to (2), if it be assumed that fundamentally there is not perfect correlation of EQ and IQ, then even in a perfectly sectioned school, geometrically it would be expected (be perfectly "normal") to have more than half of the dull people with IQ's more than 1.00 and half of the bright people with AQ's of less than 1.00. The area *yoc* diminishes in size as *r* becomes larger, but does not disappear until *r* becomes 1.00. Thus the AQ, as a statistical measure, is the old problem of trying to lift one's self by one's boot straps. In all the above it is assumed that the distribution of both IQ's and EQ's is normal, which may or may not be true.

It should be noted that, with all its defects, the AQ method *will pick out* the school or the individual who has an AQ of such magnitude that it *would seldom happen in a normal correlation* plot between EQ and IQ. Like the IQ, a difference of 0.01 in AQ in the case of two people with AQ's respectively of 0.65 and 0.66 is a different amount from the difference between two persons with AQ's of 0.99 and 1.00 respectively. Perhaps by a more complicated mathematical procedure,

wherein we calculate the probability of a child "working up to capacity," we may yet improve the measuring value of the AQ as well as secure its value as an incentive method.[1] A rough empirical beginning in this direction has been made by Pintner in setting up his $\pm$ boundry lines of "normal effort." We may reiterate the time-worn statement: The simplest explanation is by no means necessarily the truest. At best, under present circumstances, using norms computed in the usual way (norm = average score for a given age), AQ's of more than 1.00 seem just as logical and just as necessary (although at present not as abundant, as heretofore pointed out) as AQ's less than 1.00. The explanation thus suggested by the geometrical approach is that both the positive and negative differences so far found, likewise all AQ's not 1.00, are entirely due to the lack of perfect correlation between the mental and educational indices, part of the differences indicating true school maladjustment, and part being as yet undetermined.

The use of the regression of age on test score in norms will alter the proportions of the above diagram but will not eliminate AQ's of more than 1.00, if both MA's and EA's are computed by the same method.

Whether the AQ can be brought up to 1.00 or to 1.50 or to any predetermined figure, depends to a very great extent upon the EA used, the nature of the scale or examination, the norms used and the school subject under consideration. Presumably a child with good arithmetical ability might be easily brought up to an AQ of 1.00 in the "four fundamental operations" of arithmetic since his EQ, the numerator of equation (1), might be easily brought up to normal, while it might be found much more difficult to bring him up to an AQ of 1.00 in an arithmetical test dealing largely with arithmetical reasoning. We no longer talk of the "rote memory type" and "logical memory type" of person, but we know that there are apparently rare cases of "special abilities" and "disabilities" in school subjects, and even in different parts of the same school subject. Neurologists and psychologists disagree regarding their neural basis, and proper remedial treatment. Such cases of "disability" are "real enough" to the teacher to be labeled with the term, even though theoretically they may prove to be non-existent.

[1] Since writing this article, the authors have been privileged to read an as yet unpublished manuscript by J. C. Chapman wherein he demonstrates to his own satisfaction that the difference, between educational age and mental age, obtained from a single testing, is quite too unreliable for individual readjustment of pupils.

EDITOR'S NOTE: Dr. Chapman's article is published in this issue.

Other Possible Explanations of the Less than Unity Correlation Between EA and MA

Pintner's norms are based on the very pupils upon whom he reports. Pintner assumes that anyone whose educational index is more than eight points advanced over his mental index is advanced in motivation, and *vice versa*. Had he called "advanced" all people who were on the plus side, and "retarded" all who were on the minus side, then with these two groups he would find approximately 50 per cent of retarded motivation and 50 per cent of advanced motivation. In the light of an article by Toops and Pintner[1] dealing with pupils of the same city, and social status similar to that of those who determine Pintner's intelligence norms, it becomes evident that one very important reason for so many of the mentally advanced pupils being retarded in accomplishment is the fact that many bright pupils are promoted by chronological age rather than by ability to progress and so have not had the chance to come up to normal by being given opportunity to do advanced work. In the article cited it was found that only 15 per cent of the total of 1218 pupils were advanced in school one semester or more, while, by the same standard, 37 per cent were retarded in school one semester or more. With such a state of affairs, it undoubtedly is true that many dull pupils are attempting too difficult work, while it is assuredly true that many bright pupils are not attempting as difficult work as they are capable of doing.

Another factor which always operates in mental measurements is attenuation. The fact that at present the correlation between mental and educational indices is less than 1.00 may be partly explained by the inaccuracies in the measurements which always tend to attenuate, or lower, the correlations.

Undoubtedly other reasons for the less than 1.00 correlation are to be found in "special abilities or disabilities" and interest, actual laziness, etc. Other bad conditions in home and school exert their influence also. It will be found that most of the bright but less than 1.00 AQ pupils are doing *above average work in the grades they are now in*, where merely "passing" performance is acceptable. Pintner, Coy, Whipple, Coxe, and others have shown that motivation and more and better school work is the almost inevitable result when mentally advanced pupils are promoted in school to that point where they have

---

[1] Toops, H. A. and Pintner, R.: Mentality and School Progress. *Jour. of Educ. Psych.*, Vol. 10, No. 5–6, 1919, pp. 253–262.

the competition of pupils of about the same mental ability, or rates of progress.

Dr. May, as the result of a preliminary investigation of the relationships of the hours spent in study, intelligence and school marks of college students, advises the writers that in the case of his subjects there is a decided negative correlation between hours spent in study and intelligence; this certainly means that bright students, able to "get by the passing mark" with little study prefer to spend a proportionately larger amount of their time on other than study activities; consequently, by consideration alone of the very objective measure of number of hours spent in study, it is very evident that such bright pupils would be readily able to accomplish more if required to do so by being placed in a section of bright pupils where competition between the extreme members of the class would be more truly real competition.

(*Continued in January*)

THE METHOD FOR FINDING THE CORRESPONDENCE BETWEEN SCORES IN TWO TESTS

ARTHUR S. OTIS

Yonkers-on-Hudson, New York

Purpose

The statement made by the writer that "the equation of the line which most probably expresses the true relationship between x and y is $y = \frac{\sigma_y}{\sigma_x} x$" has been challenged by eminent statisticians and for that reason it has seemed desirable to publish a proof.[1]

The statement referred to the variables x and y, as two measures of the same trait. (In the particular case under discussion the trait was general mental ability.) The values x and y were subject to errors of measurement, causing them to correlate less than 1.00 with each other.

It has been contended that the regression line, $y = r_{xy}\frac{\sigma_y}{\sigma_x} x$, expresses the true relationship between x and y, and it is with the special purpose of correcting this view that the present article is written.

Method

It has seemed desirable to give the proof in two forms; first, a proof by analogy which, while not rigorous, is nevertheless believed to be vivid and suggestive, and second, a rigorous mathematical proof.

First Proof

A Hypothetical Case.—In order to bring out clearly the difference between the two lines referred to above, namely, the line whose equation is $y = r_{xy}\frac{\sigma_y}{\sigma_x} x$, which is called a regression line, and the line whose equation is $y = \frac{\sigma_y}{\sigma_x} x$, which is called in this article the relation line, let us consider a hypothetical case of two variables. Take for example the Fahrenheit and Centigrade thermometer scales. If both were

[1] This statement appeared in the article entitled The Reliability of the Binet Scale and Pedagogical Scales. *Journal of Educational Research*, September, 1921, p. 132. In that article the reliabilities of x and y were assumed to be equal.

applied to the same thermometer portions of each scale would correspond as shown below.

C.	F.
20—	—68
15—	—59
10—	—50
5—	—41
0—	—32

That is, 0°C. measures the same temperature as 32°F., 5°C. measures the same temperature as 41°F., etc.

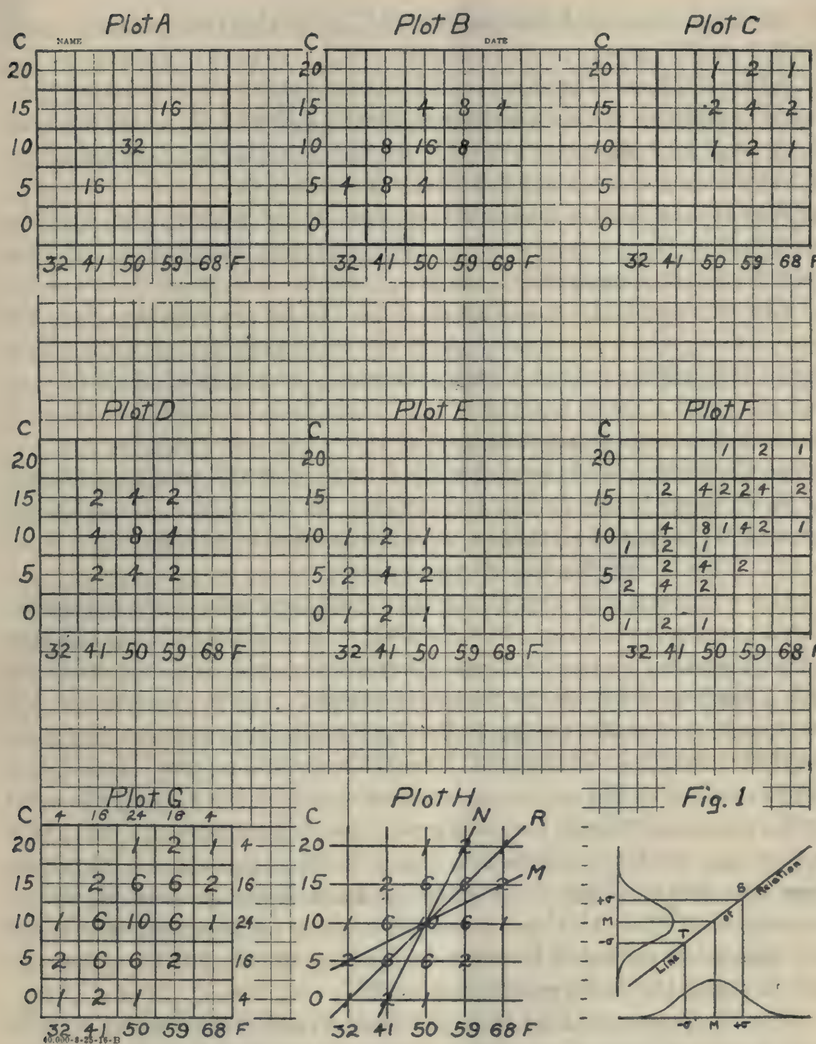
Now let us suppose that each of the several temperatures is read independently by two persons, one reading Centigrade and one Fahrenheit. Let us suppose for the moment that in a certain experiment the thermometer is read by both persons,

16 times while standing at 15°C.,
32 times while standing at 10°C., and
16 times while standing at 50°C.

If the readings by both individuals are accurate in all cases and if plotted, these would appear as shown in Plot A.

For the sake of introducing the factor of error, let us suppose, instead, that the person reading the Fahrenheit scale stands so far away from the thermometer that the numbers are indistinct so as to be often misread. Let us suppose that half the readings at each temperature are correct, that one-fourth are one graduation too high, and one-fourth are one graduation too low. If the readings of temperature thus made by the two persons were plotted these would appear as shown in Plot B.

Now let us suppose both persons were to read the thermometer from so far away, as to make similar errors, half of the readings of each temperature being correct, one-fourth too high, and one-fourth too low. This will give us the sort of correspondence between unreliable readings of the same temperature by two different scales that is found between the two unreliable measurements of mental ability by two mental ability tests. Each of the numbers 4, 8, and 4 in the 15°C. row of Plot B would in this case be split vertically into a fourth, a half, and a fourth so that the 16 readings of actual temperature 15°C., by the two persons, when plotted would appear as shown in Plot C. Similarly the 32 readings of actual temperature 10°C., by the two persons, when plotted would appear as shown in Plot D. Similarly


Plot A
Plot B
Plot C
Plot D
Plot E
Plot F
Plot G
Plot H
N
R
M
Fig. 1
Line of Relation
C
F
32 41 50 59 68 F
20
15
10
5
0

the readings of actual temperature 5°C., by the two persons, when plotted would appear as shown in Plot E.

When the pairs of readings of all 64 temperatures were plotted these would constitute the summation of Plots C, D and E, as shown combined in F and summated in G. At the top and right edges of the plot are shown the totals of the columns and rows.

Plot G has been converted into Plot H by placing the numbers representing the frequency of readings at the intersections of lines instead of in the squares.

The Regression Line.—Now in Plot H, let us consider first the four cases in which the temperature was read as 32°F. In these four cases the readings on the Centigrade scale were 1 at 0°, 2 at 5°, and 1 at 10°, with a mean reading of 5°. Next take the array[1] of 16 cases in which the temperature was read at 41°F. In this array the readings on the Centigrade scale were 2 at 0°, 6 at 5°, 6 at 10°, and 2 at 15°, the mean of these being 7.5°. And so on. If we drew a straight line through the means of all these arrays the line would be located as shown at M. This is called a *line of regression.*

Since the mean of the Centigrade readings, which are associated with Fahrenheit readings of 32°, is 5°C., it is said that 5°C. is the *most probable* reading on the Centigrade scale which *will be found associated* with a reading of 32° on the Fahrenheit scale. Or, in other words, if a 65th reading is made on the Fahrenheit scale, under the same conditions[2] and this is a reading of 32°, and it is desired to predict what will be the reading of the same temperature made on the Centigrade scale by the other individual, the best prediction is a reading of 5°C. It is in this way that the regression line is used in prognosis. Similarly since the mean of the Centigrade readings found associated with a reading of 68°F. is 15°C., it is said that given a Fahrenheit reading of 68°, the most probable Centigrade reading which will be associated with it under the same conditions is 15°C.

Why the Regression Line Does Not Show True Correspondence.—Why is it, however, that the mean Centigrade reading found associated with readings of 32°F. is 5°C. when the Centigrade value corresponding to 32°F. is known to be 0°C.? The answer is as follows: In the first place all 4 of these readings of 32°F. are in error downwards by hypo-

---

[1] The distribution of values on one scale associated with a single value on the other scale is called an *array.*

[2] The meaning of this expression will be brought out later.

thesis since the actual temperature read in each case was 41°F., as shown in Plots A and E.

Now 41°F. is the same as 5°C. so one would naturally expect the average of readings of the 4 temperatures of 5°C. to be 5°C.

Similarly the mean of the array of Centigrade readings found associated with readings of 41°F. is at 7.5°C. but this is not the same temperature as 41°F. which is only 5°C. And, as before, the explanation is that of these 16 readings of 41°F., 8 were correct readings of actual temperatures of 41°F. and 8 were incorrect readings of 50°F. The average of these 16 actual temperatures is 45.5°F. and this is the same temperature as 7.5°C. As before, one would naturally expect that the mean reading of actual temperatures averaging 7.5°C. would be 7.5°C.

The mean of the array of Centigrade readings found associated with readings of 50°F. is 10°C. and 10°C. = 50°F. This case differs from the preceding in that 50°F. happens to be the mean of all the Fahrenheit readings and consequently the mean of the 24 actual temperatures read as 50°F. was exactly 50° which equals 10°C. so naturally the mean Centigrade reading of these temperatures would be expected to be 10°C.

Going up the scale we find the mean Centigrade reading found associated with readings of 59°F. is 12.5°C. instead of 15°C. which equals 59°F. and we find the mean Centigrade reading found associated with readings of 68°F. is 15°C., whereas 68°F. corresponds to 20°C. The chief point to be noted in this connection, however, is that if we did not know in advance what number of degrees Centigrade denoted the same temperature as 32°F. we could not find it by taking the mean of the array of Centigrade readings found associated with readings of 32°F. for the obvious reason that the number of degrees Centigrade denoting the same temperature as 32°F. is 0° while the mean Centigrade reading found associated with readings of 32°F. is 5°C.

The same is true all the way up the scales with the single exception of 50°F. in this particular case, because it is the mean of all the Fahrenheit readings. The procedure which should be adopted to find the Centigrade reading corresponding to any given Fahrenheit reading will be described later.

The Meaning of Regression.—It will be seen that instead of the means of the arrays of Centigrade readings found associated with each of the Fahrenheit readings

32°, 41, 50°, 59°, and 68° being respectfully

0°, 5°, 10°, 15°, and 20°C to correspond, they were in reality 5°, 7.50°, 10°, 12.5° and 15°C.

The mean value of Centigrade readings found associated with each of the Fahrenheit readings tend to be nearer to the mean (10°) of all the Centigrade readings than are the Centigrade values to which these Fahrenheit readings correspond.

It is said that the means of these arrays of Centigrade readings *regress* (fall back) toward the mean (10°) of all the Centigrade readings. That is why the line is called a "regression line."

There are Two Regression Lines.—In the same way it may be seen that the means of the arrays of the Fahrenheit readings corresponding to the several Centigrade readings regress toward the mean (50°) of these Fahrenheit readings so that if a line is drawn in a plot through these means it will take the position shown at N in the Plot H. This is the other regression line, there being two in every such case, one through the mean of the vertical arrays and one through the mean of the horizontal arrays.

A Generalization.—We may now make a very general statement and say that whenever x and y values are plotted and do not correlate perfectly, the mean of every array of y values associated with any single value of x is *nearer to the mean* of all the y values than is the value of y which truly corresponds to that single value of x.

Effect of Shifting Distributions.—It should be noted that while the true value of the temperatures read as 68°F. was in this particular case 20°C., nevertheless if the 16, 32 and 16 temperatures had been at 50°, 59°, and 68°F. respectively the mean of the true values of the temperature then read as 68°F. would have been 17.5°C. And if the 64 temperatures had been at 59°, 68° and 77°F. the mean of the true values of temperatures then read as 68° would have been 15°C. This means that if the regression line were used in the effort to determine the true Centigrade value corresponding to 68°F., this would be found to be 20°C., in one case, 17.5° C. in another, and 15° in another.

The value of one variable which will most probably be found associated with a given value of the other variable varies therefore according to the general position of the values investigated on the scales.

Use of the Regression Line in Mental Testing.—Now let us see the significance of this statement as applied to mental measurement. Suppose we have tested a group of Grade XII pupils with Forms A and B of a Mental Ability Test, and wish to find the most probable score a pupil will have made (or will make) in Form B who has made a

score of 100 in Form A. This is done by means of the regression line which indicates the theoretical mean of the B scores found associated with an A score of 100. Although the B score truly corresponding to the A score of 100 might be also 100, the mean of the associated B scores might be 110, showing that a pupil in this group making a score of 100 in Form A would most probably have made a score of 110 in Form *B*. This is because 100, being a low score for such a group, is most probably in error downwards. And it may be said also in the case of any other Grade XII pupil who has taken Form A only, but has made 100 points, that insofar as he is typical of the Grade XII pupils of the group considered, he, too, will most probably make a score of 110 in Form B. This merely amounts to saying that if a typical Grade XII pupil makes a score of 100 points in this test, his score is most probably in error by 10 points downward, and that this error tends to be corrected in his second score.

On the other hand, if a group of Grade V pupils were tested with the same two forms, A and B, then by means of the regression line in the new plot it might be found that the mean of the B scores found associated with A scores of 100, was only 90, showing that a Grade V child who made a score of 100 in Form A will most probably have made a score of 90 in Form B. This is because a score of 100, being for a fifth grader a *high* score is most probably in error *upwards*. And of any other Grade V pupil who has made a score of 100 in Form A it may be said that if he is typical of the fifth graders who took both forms, he too will most probably make a score of 90 in Form B. This merely amounts to saying that if a Grade V pupil makes a score of 100 the probability is that his score is in error by 10 points upward, and that in a second score this error tends to be corrected.

The regression line therefore shows the most probable *true* score in a second test which a pupil would obtain who made a given score in a first test, the score in the first test being *in error*. The regression line therefore does *not* show the true correspondence between true scores in both scales.

How May the True Correspondence be Found?—We come now to the problem of finding the true line of relation between two variables when we have before us only the plot such as Plot *G* showing the incomplete correspondence between the two variables.

Let us go back to plot *A* and trace the evolution of the standard deviations of the two variables. In plot *A*, σ_F (the standard deviation of the 64 F. readings) $= 9\sqrt{\frac{1}{2}}$ and σ_C (the standard deviation of

the 64 C. readings) $= 5\sqrt{\frac{1}{2}}$. Next we assumed that errors which occurred in the Fahrenheit readings were distributed thus:

Errors	−9,	0,	+9
Frequency . . .	16,	32,	16

Here, σ_{eF} (the standard deviation of the errors of Fahrenheit readings) $= 9\sqrt{\frac{1}{2}}$.[1] We assumed also that errors which occurred in Centigrade readings were distributed thus:

Errors	−5,	0,	+5
Frequency . . .	16	32	16

Here, σ_{eF} (the standard deviation of errors of Centigrade readings) $= 5\sqrt{\frac{1}{2}}$.

Variabilities of Observed Measures are Proportional to Variabilities of True Measures.—It will be seen that the magnitudes of the errors made in the two scales (as measured by their standard deviations σ_{eF} and σ_{eC}) have the same ratio (9:5) as the standard deviations of the true temperatures themselves in the two scales. That is: $\sigma_{eF} : \sigma_{eC} : \sigma_F : \sigma_C$. This is for the obvious reason that an error of 9 degrees on the Fahrenheit scale equals an error of 5 degrees on the Centigrade scale. The effect of these errors is such therefore that the standard deviation of the *observed* measures on the Fahrenheit scale is $\frac{9}{5}$ of the standard deviation of the observed measures on the Centigrade scale. Or, to put it the other way round, the *ratio of the standard deviations of the true Fahrenheit and Centigrade measures is the same as the ratio of the standard deviations of the observed Fahrenheit and Centigrade measures which is as* 9:5.

The Correspondence between Means.—As has been shown, the mean of the whole distribution of values of either variable does not tend to be in error either upward or downward and therefore the mean of the whole distribution of values of one variable probably truly corresponds to the mean of the whole distribution of values of the other variable.

The Relation Line.—Going back to Plot H, then, if we wish to find the *true correspondence* between Fahrenheit and Centigrade values, we must draw a line through the point representing the mean (50) of all the Fahrenheit readings and the mean (10) of all the Centigrade readings, such that for every 9 units on the horizontal scale the line rises 5 units on the vertical scale. This is the line R. The line R, then, expresses the *true relation* between the Fahrenheit and Centigrade scales and is called the Relation Line.

[1] There is no necessary connection between this and σ_F).

A Further Generalization.—We may now make the general statement that whenever two measures, x and y, of the same trait (such as scores in two tests of the same ability) are not perfectly correlated, and there is no evidence that one test is any more reliable than the other, *the line which most probably represents the true relationship between the two measures is the line* $y = \frac{\sigma_y}{\sigma_x} x$ when the means of the values of the variables have been taken as the zero points from which to measure the variables. This is the line drawn through the point representing the means of the two groups of measures and through the point S representing $+ 1\sigma$ in each distribution and through the point T representing $- 1\sigma$ in each distribution, as shown in Fig. 1. Stated in other words, the true correspondence between such measures is probably such that the mean of the measures of one variable equals the mean of the measures of the other variable, and the standard deviation of the observed values of one variable represents the same increment of ability as the standard deviation of the observed values of the other variable.

In this proof there is an underlying assumption throughout that the two scales by which the variables are measured are so constructed that the relationship is rectilinear, by which is meant that the units of one scale bear a constant relation to the units of the other scale throughout, so that the true line of relation is a straight line.

Cases in which the line of relation are not straight must be dealt with as discussed on page 125 of the article referred to and also in the Reliability of Spelling Scales, *School and Society*, October 28–November 18, 1916.

Second Proof

Hypothesis.—Let us suppose we have two mental ability tests, X and Y.

Let X_1, X_2, X_3, etc., represent the scores obtained in Test X by the different individuals, and Y_1, Y_2, Y_3, etc., represent the scores obtained by the same individuals in Test Y. Thus, X without a subscript represents any score obtained in Test X, and Y represents any score obtained in Test Y.

Let x_1 represent the mean of a very large number of scores of the first individual in Test X and be considered, therefore, as the true score of that individual in Test X. Let x_2, x_3, etc., represent similarly the true scores of the other individuals in Test X.

Let $X_1 - x_1 = e_1$, $X_2 - x_2 = e_2$, etc. The value e, therefore, is the amount by which the obtained score of any individual differs from his true score as defined. Similarly, let $Y_1 - y_1 = f_1$, $Y_2 - y_2 = f_2$, etc. Generally speaking then $X - x = e$ and $Y - y = f$.

The variables e and f may be considered as errors of measurement,[1] and obviously they are totally uncorrelated with each other and with x and y.

For the sake of simplicity let us assume that the values of X, x, Y, and y, are measured from their respective means so that

$$\Sigma X = 0, \sqrt{\frac{\Sigma X^2}{N}} = \sigma_x,$$

and the same for x, Y, and y.

The quantities, e and f, will be sometimes positive and sometimes negative and we may assume them to be distributed normally in each case with the mean at zero, in which case the mean of the X values is the same point on the X scale as the mean of the x values, and the same for Y and y.

While we have spoken of Tests X and Y as both being mental ability tests, it is not certain, of course, that the traits measured by the two tests are absolutely identical. In other words, r_{xy}, the correlation between what we have called true scores in Test X and true scores in Test Y, may be slightly less than +1.00. But, for the time

[1] There are, of course, influences affecting scores in a mental ability test, such as varying degrees of effort, etc., which are theoretically distinguished from mental ability itself but which nevertheless may be correlated, either positively or negatively, with mental ability as defined. Thus it is conceivable that dull pupils might try harder to score well in a mental ability test than bright pupils, so that effort might correlate negatively with mental ability in a certain group. But in so far as effort *is* correlated either one way or the other with mental ability, just to that extent the test score measures effort (or the opposite of effort) as well as mental ability and equality of effort will tend to make for equality of score in the same way that equality of mental ability does, although, of course, to a lesser extent. In other words, mental ability as measured is not mental ability as defined, and when we speak of the reliability of a test, we mean the consistency of its scores—the degree to which two scores of the same individual correspond. In that sense all factors which contribute consistently to the score and thereby operate to make the two scores of the same individual in the same test equal are to all practical purposes part of the ability tested, and we may as well consider the effects of those factors which cause two scores of the same individual in the same test to differ as being to all practical purposes errors of measurement. This is not essential to the proof, however.

being, let us assume that $r_{xy} = +1.00$ and later we will consider the case in which $r_{xy} < +1.00$.

Let us suppose it is desired to find the most probable relation between true values of x and true values of y. In other words, let us suppose it is desired to find the relation between two values, x and y, when these measure the same amount of the trait.

It will now be shown[1] that this relation is expressed by the equation

$$y = \sqrt{\frac{r_{YY}}{r_{XX}}}\frac{\sigma_Y}{\sigma_X}x \tag{A}$$

in which r_{XX} is the reliability coefficient of variable X, and r_{YY} is the reliability coefficient of variable Y.

Proof.—Assuming that $r_{xy} = +1.00$, let $y_1 = mx_1$, $y_2 = mx_2$, $y_3 = mx_3$, etc. The constant, m, is the ratio, therefore, of the units of the two scales; and the tangent of the angle of the line which represents the true correspondence between measures of the two scales is therefore equal to m.

If
$$y = mx \tag{1}$$
then
$$y^2 = m^2x^2 \tag{2}$$
$$\Sigma y^2 = m^2\Sigma x^2 \tag{3}$$
$$\sigma^2_y = m^2\sigma^2_x \tag{4}$$
$$\frac{\sigma^2_y}{\sigma^2_x} = m^2 \tag{5}$$
$$m = \frac{\sigma_y}{\sigma_x} \tag{6}$$

Of course, we do not know the value of σ_y and σ_x because these are standard deviations of true scores which we cannot obtain but it will be shown now how to find the value of $\frac{\sigma_y}{\sigma_x}$ from the values of σ_X, σ_Y, r_{XX}, and r_{YY}, which can be found.

By definition,
$$X = x + e \tag{7}$$
Squaring,
$$X^2 = x^2 + 2ex + e^2 \tag{8}$$
Summating,
$$\Sigma X^2 = \Sigma x^2 + 2\Sigma ex + \Sigma e^2 \tag{9}$$
Now by the formula for correlation,
$$r_{ex} = \frac{\Sigma ex}{\sqrt{\Sigma e^2 \Sigma x^2}} \tag{10}$$

---

[1] It should be borne clearly in mind that it is *not* sought to prove that the equation is to be used to find the most probable value of Y that will be found associated with a given value of X, nor that it is to be used to find the most probable true measure, in terms of a Y scale, of the trait in an individual who has attained a given measure, X, in another scale. This formula is not to be used for prediction or for estimating true values in one scale from obtained values in another. For these purposes the regression equation should be used.

But by hypothesis, $$r_{ex} = 0 \tag{11}$$
Therefore, $$\Sigma ex = 0 \tag{12}$$
From equations 9 and 12, $$\Sigma X^2 = \Sigma x^2 + \Sigma e^2 \tag{13}$$
Whence, $$\sigma^2_X = \sigma^2_x + \sigma^2_e \tag{14}$$
or $$\sigma^2_x = \sigma^2_X - \sigma^2_e \tag{15}$$

Equation 14 shows that the standard deviation of a distribution of true scores is augmented by the introduction of errors to the extent of the standard deviation of the distribution of errors.

Now by a formula[1] devised by the writer,

$$r_{XX} = 1 - \frac{\sigma^2_e}{\sigma^2_X} \tag{16}$$

in which r_{XX} is the reliability coefficient of correlation between scores in Test X, and e has the same meaning as used above.

Now by equation 16, $$r_{XX}\sigma^2_X = \sigma^2_X - \sigma^2_e \tag{17}$$
By equation 15, $$\sigma^2_x = \sigma^2_X - \sigma^2_e \tag{18}$$
Therefore, $$\sigma^2_x = r_{XX}\sigma^2_X \tag{19}$$
and $$\sigma_x = \sqrt{r_{XX}}\sigma_X \tag{20}$$

This equation constitutes a formula for finding the standard deviation of true scores of a group of individuals from the standard deviation of the obtained scores of those individuals, knowing the reliability coefficient of correlation obtained from the same group of individuals.

Similarly $$\sigma_y = \sqrt{r_{YY}}\sigma_Y \tag{21}$$

Therefore, $$\frac{\sigma_y}{\sigma_x} = \sqrt{\frac{r_{YY}}{r_{XX}}}\frac{\sigma_Y}{\sigma_X} \tag{22}$$

Now the equation of the line which represents the true correspondence between scores in Tests X and Y, as shown in equation 6, is

$$y = \frac{\sigma_y}{\sigma_x} x \tag{23}$$

By Equation 22 this equation becomes $$y = \sqrt{\frac{r_{YY}}{r_{XX}}}\frac{\sigma_Y}{\sigma_X} x \tag{24}$$

This then is the equation of the line which represents the true correspondence between scores in Tests X and Y, assuming that true scores in these two tests measure identical traits.

[1] This is the same formula as equation 1, page 140 of the article entitled, The Reliability of the Binet Scale and Pedagogical Scales, *Journal of Educational Research*, September, 1921.

The Correspondence between Two Forms of a Test.—Now if we are dealing with two "forms" of the same test, the presumption is that one form is just as reliable as the other, in which case we may assume that $r_{XX} = r_{YY}$ and hence $\sqrt{\frac{r_{YY}}{r_{XX}}} = 1$

It is reasonable to assume also that the correlation between true scores in the two forms is practically perfect, that is, the two forms may be assumed to measure identical traits, so we may call r_{xy} equal to $+1.00$. In this case therefore, Equation 24 becomes simplified, so that the equation[1] of the line which most probably represents the true correspondence between the scores of the two forms of a test is

$$y = \frac{\sigma_Y}{\sigma_X} x. \tag{25}$$

The way Equation 25 is used is as follows: Suppose it is desired to find the correspondence between scores in Form A of the Otis Higher Examination given as an initial test and Form B of the same examination given a week later, so that scores in Form B, so given, could be transmuted into terms of Form A, so given, for comparative purposes. Both forms would be given to the same group of individuals, Form A first and Form B a week later. Let us suppose the mean of the Form A scores is found to be 50 points and the mean of the Form B scores to be 52 points. Let us suppose σ_A, the standard deviation of the scores in Form A, is found to be 11 points, and σ_B, 10 points. We would then assume that 50 points in Form A, so given, corresponds to 52 points in Form B, so given, and that measuring the scores from their respective means, any score in Form B equals $\frac{10}{11}$ the corresponding score in Form A.[2]

The Case in Which Tests Do Not Measure Identical Traits.—Now let us consider the case in which $r_{xy} < +1.00$, that is, the case in which the

[1] When we are considering the correspondence between scores, we are referring of course to true scores. When we say, for example, that 50°F. corresponds to 10°C., we mean of course that a true temperature of 50°F. corresponds to a true temperature of 10°C., not that some temperature erroneously read as 50°F. corresponds to some temperature erroneously read to 10°C. Similarly, when we speak of the correspondence between scores in Tests X and Y we refer to the correspondence between true scores, x and y. For that reason the equation of the line is given in a form expressing the correspondence between true scores, x and y, in terms of obtained scores, X and Y.

[2] This method is suitable, of course, only in case it is assumed that the relationship is rectilinear

true score of an individual in Test X does not measure exactly the same combination of traits as the true score of the individual in Test Y. In what sense, then, may there be a true correspondence between scores in Tests X and Y? It would seem that there can be a true correspondence only with respect to the measurement of that trait or group of traits which is measured by both tests.

Now the true score x of any individual in Test X, as defined above, will differ slightly from the true score that he would obtain in Test X if the effect of certain factors specific to Test X were cancelled so that the score in Test X was affected only by factors which affected a score in Test Y also.

Let this difference in score in Test X be represented by s.

Let a similar difference in score in Test Y be represented by t.

Let g represent the true score (average of a large number of scores) of an individual in Test X when the effect, s, of factors specific to Test X are cancelled; that is, when $\sigma_s = 0$. According to these definitions,

$$x = g + s \tag{26}$$

Let
$$y = h + t \tag{27}$$

From Equation 26,
$$x^2 = g^2 + 2gs + s^2 \tag{28}$$

and
$$\Sigma x^2 = \Sigma g^2 + 2\Sigma gs + \Sigma s^2 \tag{29}$$

but
$$r_{gs} = \frac{\Sigma gs}{\sqrt{\Sigma g^2 \Sigma s^2}} = 0^1 \tag{30}$$

whence
$$\Sigma gs = 0 \tag{31}$$

Therefore
$$\Sigma x^2 = \Sigma g^2 + \Sigma s^2 \tag{32}$$

and
$$\sigma^2_x = \sigma^2_g + \sigma^2_s \tag{33}$$

or
$$\sigma^2_g = \sigma^2_x - \sigma^2_s \tag{34}$$

Similarly,
$$\sigma^2_h = \sigma^2_y - \sigma^2_t \tag{35}$$

Now, as in Equation 16,
$$r_{xy} = 1 - \frac{\sigma^2_s}{\sigma^2_x} \tag{36}$$

Multiplying by σ_x^2,
$$r_{xy}\sigma^2_x = \sigma^2_x - \sigma^2_s \tag{37}$$

Now by Equations 34 and 37,
$$\sigma^2_g = r_{xy}\sigma^2_x \tag{38}$$

Similarly,
$$\sigma^2_h = r_{xy}\sigma^2_y \tag{39}$$

Therefore,
$$\frac{\sigma^2_h}{\sigma^2_g} = \frac{\sigma^2_y}{\sigma^2_x} \tag{40}$$

and
$$\frac{\sigma_h}{\sigma_g} = \frac{\sigma_y}{\sigma_x} \tag{41}$$

This equation shows that the ratio of the standard deviations of the true scores in Tests X and Y (true scores being now defined as scores in

---

[1] Since s factors are specific to x by hypothesis, therefore $r_{sh} = 0$. And by hypothesis $r_{gh} = +1.00$. Therefore $r_{gs} = 0$.

which the effect of all factors not common to both tests have been neutralized) is equal to the ratio of the standard deviations of the true scores as previously defined.

Now the true scores (g and h) in Tests X and Y (as measures of the same trait) are of course perfectly correlated so that each value of h is some constant times the corresponding value of g. Let us represent this constant by m.

Then
$$h = mg \tag{42}$$
$$h^2 = m^2g^2 \tag{43}$$
$$\Sigma h^2 = m^2\Sigma g^2 \tag{44}$$
$$m^2 = \frac{\Sigma h^2}{\Sigma g^2} \tag{45}$$
and
$$m = \frac{\sigma_h}{\sigma_g} \tag{46}$$

The value of m is by definition the tangent of the angle of the line of true correspondence between scores in Tests X and Y as measures of the same trait. The equation of the line is therefore

$$h = \frac{\sigma_h}{\sigma_g} g \tag{47}$$

Substituting in this equation the value of $\frac{\sigma_h}{\sigma_g}$ found in Equation 41, the equation of the line becomes

$$h = \frac{\sigma_y}{\sigma_x} g \tag{48}$$

Substituting in this equation the value of $\frac{\sigma_x}{\sigma_v}$ found in Equation 22, the equation of the line becomes

$$h = \sqrt{\frac{r_{YY}}{r_{XX}}} \frac{\sigma_Y}{\sigma_X} g \tag{49}$$

We might as well do away with the ultrafine distinction, however, between g and x and between h and y and let x and y represent the true scores in Tests X and Y as measures of the same trait, thereby getting back to familiar symbols. In that case Equation 49 becomes

$$y = \sqrt{\frac{r_{YY}}{r_{XX}}} \frac{\sigma_Y}{\sigma_X} x \tag{50}$$

in which the values of all the variables are measured, of course, from their respective means.

Application of the Formula.—Equation 50 would be used in the following way: Suppose it is desired to find the true correspondence

between scores in the Binet Scale and the Otis Higher Examination. Call these Tests X and Y. Suppose these tests to have been administered to the same group of individuals. Suppose the standard deviations (σ_x and σ_y) of scores in the two tests by this group are 15 and 18 respectively and suppose the reliability coefficients of correlation (r_{XX} and r_{YY}) obtained with this same group[1] to be 0.90 and 0.80 respectively. The correspondence between scores will be expressed by the following equation:

$$\text{Otis score (measured from mean)} = \sqrt{\frac{80}{90}} \times \frac{18}{15} \times \text{Binet Score (measured from mean)} \qquad (51)$$

If the reliabilities of the two tests are not known or for other reason are considered as equal, Equation 50 becomes, of course, simply:

$$y = \frac{\sigma_y}{\sigma_x} x \qquad (52)$$

Derivation of the Regression Equation.—Now suppose variable X is a measure of age or some quantity not subject to errors of measurement so that we may call r_{XX} equal to $+1.00$. Then the correspondence between X and Y (Equation 50) becomes:

$$y = \sqrt{r_{YY}} \frac{\sigma_Y}{\sigma_X} x \qquad (53)$$

Now it may be shown that if $r_{XX} = +1.00$, $\sqrt{r_{YY}} = r_{XY}$. Equation 53 then becomes

$$y = r_{XY} \frac{\sigma_Y}{\sigma_X} x \qquad (54)$$

This, of course, is the regular regression equation, showing that to find the score corresponding to (or normal for) any age, we may use the line of regression, that is, the line passing through the central tendencies of the arrays of scores for the several ages.

A Correction

In the May, 1922, number of this journal there appeared an article by the writer entitled, A Method of Inferring a Change in a Coefficient

[1] If the reliability coefficient of correlation for either test has been determined using a group of a different heterogeneity from the present group it will be necessary to correct the coefficient for this difference in heterogeneity by a method explained in an article by the writer entitled, A Method of Inferring the Change in a Coefficient of Correlation Resulting from a Change in the Heterogeneity of the Group, *Journal of Educational Psychology*, May, 1922. (See correction below.)

of Correlation Resulting from a Change in the Heterogeneity of the Group. In this article the last equation (not numbered) is an error. This equation should read:

$$r'_{xy} = 1 - (1 - r_{xy}) \frac{\sigma^2_x}{\sigma^2_x{}'}$$

in which r'_{xy} and $\sigma^2_x{}'$ refer to one degree of heterogeneity of the group and r_{xy} and σ^2_x refer to the other degree of heterogeneity of the group.

The application of this method is as follows.

Suppose r_{xy}, the correlation between Forms A and B of a test in Grade VI, is 0.75.

Suppose r'_{xy}, the correlation between Forms A and B of the same test in a group combining Grades IV, V, VI, VII, and VIII, is sought.

Suppose σ_x, the standard deviation of scores in Form A in Grade VI, is 40 points.

Suppose σ_x', the standard deviation of scores in Form A in the group combining the five grades, is 50 points.

Then $$r'_{xy} = 1 - (1 - 0.75) \frac{40^2}{50^2}$$

$$r'_{xy} = 0.84.$$

It may be remembered simply that the deviation of the coefficient from unity varies inversely as the square of the variability of the group.

THE LIMITS SET TO EDUCATIONAL ACHIEVEMENT BY LIMITED INTELLIGENCE

MARGARET V. COBB

Institute of Educational Research

Teachers College

(*Concluded from November*)

IV. Intelligence and Progress in High School Subjects.—We can now turn to our data on specific high school subjects, and ask what intelligence is required for success in each. In general, the figures show that the present courses in algebra make as great a demand on intelligence as do those in any one subject taken in the freshman year, and this will be reported as a sample study. More failures occur; more pupils drop out. There is a correspondence between the score a pupil makes on a general intelligence test, and the probability of his "passing" the course. There is a closer correspondence between his score on a test designed to measure mathematical ability, or ability to learn algebra, and the probability of his passing the course. It is worth while, if vocational or educational guidance is to be given, or if sections are to be made up on the basis of probable progress, to have both of these measures.

Pupils who elect algebra, or who choose a course including algebra, are in general a more intelligent group than those who do not; pupils who pass in algebra are in general a more intelligent group than those who do take it but fail. The groups overlap considerably, but the one is definitely better than the other. The graphs in Figs. 5 to 10 show the distribution of Alpha scores of pupils passing in algebra, of those who fail, and of those who do not take algebra. The contrast between the median scores of those who "pass" algebra and those who fail, or do not take it, is quite striking. In Alma, the median Alpha score of freshmen who passed in algebra was 94, while the median of those who failed was 78. In Mt. Clemens the corresponding medians were, for those who passed algebra, 107; for those who failed in algebra, 89, and, still more significant, for those who did not take algebra, 69. In Mt. Pleasant the median Alpha score of the pupils who passed algebra was 89; of those who failed, 65. In Milan, the median Alpha score of the pupils who passed was 86, while that of those who failed was 75. In Detroit (Terman Group Test of Mental Ability), the median score of those who passed was 94 and of those who failed 84.

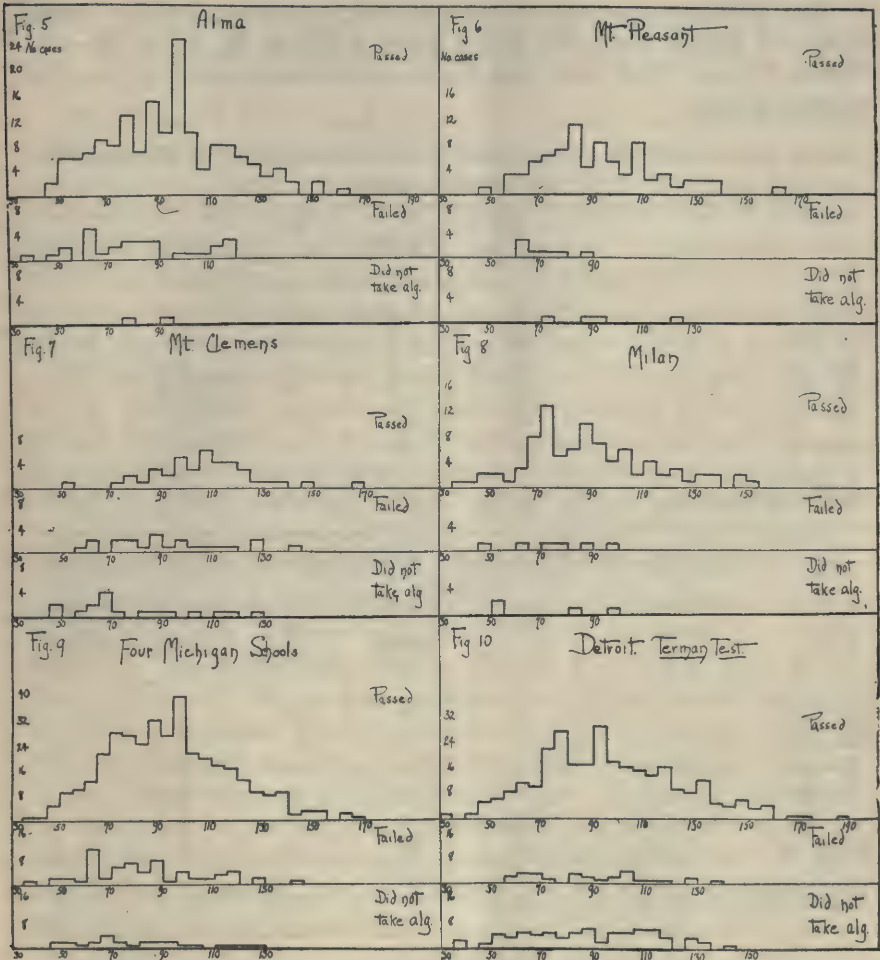


Fig. 5.—Distribution of Alpha Scores of pupils who passed, failed, or did not take algebra. Alma, Michigan.

Fig. 6.—Distribution of Alpha Scores of pupils who passed, failed, or did not take algebra. Mt. Pleasant, Michigan.

Fig. 7.—Distribution of Apha Scores of pupils who passed, failed, or did not take algebra. Mt. Clemens, Michigan.

Fig. 8.—Distribution of Alpha Scores of pupils who passed, failed, or did not take algebra. Milan, Michigan.

Fig. 9.—Distribution of Alpha Scores of pupils who passed, failed, or did not take algebra. Four Michigan schools.

Fig. 10.—Distribution of Scores on the Terman Group Test of Mental Ability of pupils who passed, failed, or did not take algebra. Detroit, Michigan.

Our Michigan data have been analyzed to indicate also the expectation of failure when the Alpha score is below 55, 55 to 74, etc. Tables XII and XIII show this, for the schools separately, and for the combined data.

TABLE XII.—PER CENT OF FRESHMEN TAKING ALGEBRA AT EACH LEVEL WHO FAILED IN ALGEBRA

	Mt. Clemens	Milan	Mt. Pleasant	Alma	Total
135+	25	0	0	0	5
115–134	25	0	0	12	11
95–114	23	6	0	9.6	10
75– 94	47	7	6	16.7	14
55– 74	67	7	23	21	20
Below 55	0	14	0	33	20
Median Alpha score	89.2	75	67.5	79	79.6

TABLE XIII.—PER CENT OF FRESHMEN AT EACH LEVEL WHO DID NOT TAKE ALGEBRA

	Mt. Clemens	Milan	Mt. Pleasant	Alma	Total
135+	0	0	0	0	0
115–134	14	0	11	0	5
95–114	8	6	0	0	2
75– 94	17	3	6	3.6	6
55– 74	25	0	4	0	8
Below 55	67	22	0	0	16
Median Alpha score	69.4	67.5	87.5	85	77.5

Another way in which to look at this relationship is through the correlation of algebra marks with Alpha scores. This varies very much from school to school, as it does with other school subjects, according to the content and method of the course, the skill of the teacher in motivating and in teaching both dull and bright pupils, in judging of their acquirements and their progress, and in assigning marks in keeping with these. Were these at their highest, and the Alpha examination a "perfect" measure of "general intelligence" the correlation would be closer—though never 1, since Alpha even then would doubtless be far from an exact measure of the specialized type of

intelligence for which algebra calls. The coefficients actually obtained from the Michigan data vary from +0.15 to +0.47, centering around +0.35.

In the course of this work a number of persons familiar with high school classes in algebra have been asked to estimate in terms of intelligence quotient the degree of intelligence necessary to complete freshman algebra successfully. The various estimates run as follows: 110, 110, 110, 105 to 110, 110 (this last was for an accelerated Grade VIII class in algebra). An intelligence quotient of 110 on the Stanford Revision of the Binet scale, at the age of 14 years (that is to say, a mental age of 15-5) corresponds to an alpha score of almost 100 (98.5).

It would seem a safe conclusion that a pupil who scores from 100 to 110 (or better) on the army Alpha examination should be fairly sure of the possibility of success in the usual course in algebra, as at present taught in academic high schools. This means a mental age of 15-6 to 16-2, and, if the child begins algebra at 14, an IQ of 110 to 115. Below this, success becomes increasingly doubtful. For success in a high school course in which the subjects were for the most part definitely less difficult than algebra, these figures might be lowered by from 10 to 15 points. Proctor mentions 95 as a minimum IQ (Alpha score 67). Probably in 90 cases out of 100, it is unwse to guide the average or less intelligent than average child into the present academic high school. Unless his IQ is over 100, or his mental age definitely over 14, he should be encouraged to try some other type of training.

Additional evidence on the school progress of children who make low scores on Alpha may be gained from the comments made by Detroit high schools on their seniors who scored less than 85. We give these without omission though it should be noted that there were some cards which bore no comment.

Score	Comment
82	Industrious, but little ability.
75	Fair record. Industrious.
55	Very slow at studies, but capable in administrative work; fine character; studying nursing now.
62	A colored girl, faithful, good typist, fair in Domestic Science.
79	Lack of application. Too much interested in Girl Scout work; that her sole interest. Entered junior college.
59	Peculiar case; she did not trust herself. Often depended on others.
72	Very slow, very timid, very faithful and plodding.
79	Not the brightest, but made of good stuff. Making good in bank.
57	Sub-normal.
78	Probably never developed her powers. Calm and easy-going disposition.

75 Low ideal. Under group influence. Did not study out of school till fourth year. Now in junior college.
68 Lack of work.
78 Was not interested in school.
73 A fair student.
81 Dull.
61 Weak in English.
84 Generally weak. A cripple.

V. Geographical Differences in Intelligence.—In any application of these findings concerning our intelligence, and the proportion of us to whom an academic high-school education or, for instance, the study of algebra is an advantage, it must be remembered that smaller parts of the country—sections, states, communities—depart widely from these general figures. How unbelievably large these geographical differences may be, between states and even whole sections of the United States, may be illustrated from the data concerning the draft which appears in Volume XV of the Memoirs of the National Academy of Sciences, and data about medical officers in Bulletin 8 of the National Research Council. This information is not highly accurate, especially for the draft. The scores are for those recruits only who were examined by means of the Alpha examination, that is, those who were considered to be adequately measured by it, or "literate." Since the literacy standard was not identical in different camps, and since in a few cases it had been impossible to reexamine recruits who should have been reexamined by a non-verbal or individual method, and they were improperly included in this "Alpha only" group, there are known to be inaccuracies in the data. It is known, for instance, that the poor showing of the New Jersey recruits is at least partly due to this cause.

These inaccuracies undoubtedly compensate one another to some extent when the states are grouped, and larger areas are considered. For this purpose the following grouping has been used:

Northeast	Atlantic	Southern
Maine	New Jersey	Georgia
New Hampshire	Pennsylvania	Florida
Vermont	Delaware	Alabama
Massachusetts	Maryland	Mississippi
Rhode Island	Virginia	Louisiana
Connecticut	West Virginia	Arkansas
New York	District of Columbia	Oklahoma
		Texas
		New Mexico

South Central	North Central	Central
North Carolina	Ohio	Indiana
South Carolina	Michigan	Illinois
Kentucky	Minnesota	Iowa
Tennessee	Wisconsin	Missouri
	North Dakota	Kansas
	South Dakota	Nebraska

Western

Oregon
Washington
Montana
Idaho
Wyoming
California
Nevada
Utah
Arizona
Colorado

Table XIV

Section	Draft			Medical officers		
	Median	Median lowest state	Median highest state	Median	Lowest state[1]	Highest state[1]
I Northeast	67	62	74	139	133	144
II Atlantic	60	49[2]	66	126	118	132
III Southern	47	41	60	115	108	125
IV South Central	45	42	47	102	98	104
V North Central	62	57	64	135	132	139
VI Central	62	56	66	126	123	133
VII Western	75	64	80	140	137	144

Table XIV gives the median score for each section, for the draft and for medical officers. In order to show how closely the median for the section is representative of the states included in the group, the median

[1] Medians derived from 10 cases or fewer are here disregarded.

[2] This is New Jersey and is undoubtedly too low because of inclusion of illiterates, as previously explained.

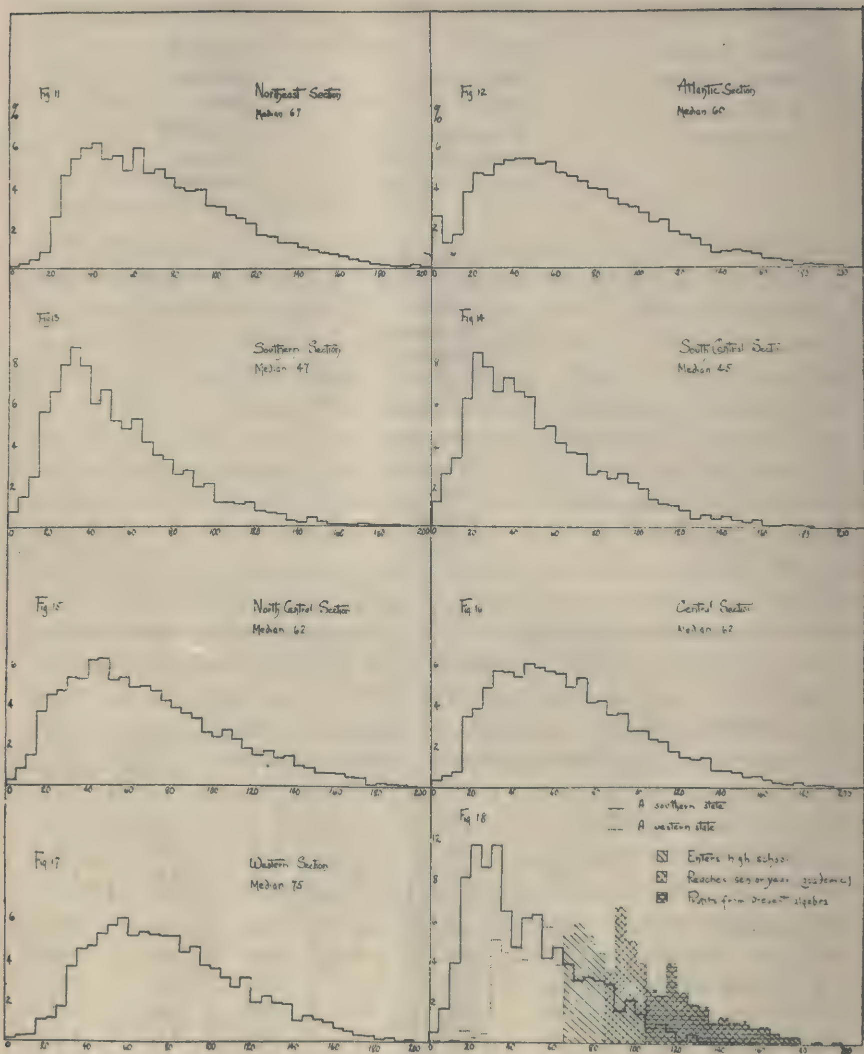


FIG. 11.—Distribution of Alpha Scores of recruits from the northeast section.
FIG. 12.—Distribution of Alpha Scores of recruits from the Atlantic section.
FIG. 13.—Distribution of Alpha Scores of recruits from the southern section.
FIG. 14.—Distribution of Alpha Scores of recruits from the south central section.
FIG. 15.—Distribution of Alpha Scores of recruits from the north central section.
FIG. 16.—Distribution of Alpha Scores of recruits from the central section.
FIG. 17.—Distribution of Alpha Scores of recruits from the western section.
FIG. 18.—Comparison of Alpha Scores of a southern and a western state.

for the lowest state and for the highest state in the group also are given.

The distributions from which these medians are derived are pictured in Figs. 11 to 17.

The evidence from the draft (40,530 cases) and from the group of medical officers (2507 cases), is in quite close agreement in indicating striking differences between different parts of the country; the rank order, and with one or two exceptions the relative size of the differences, also correspond closely.

It is obvious that these large differences in the intelligence of the population in different states has very important implications for education. Consider the comparison indicated in Fig. 18, which shows the distribution of scores for one southern (A) and one western (B) state. Suppose we make the very probable assumption that a negligible number of persons who would as adults score less than 65 on Alpha will ever as children enter an academic high school. This means, in state A, that not over 25 per cent of the population needs to be provided for in freshman classes of such high schools; but in state B, 64 per cent should be accommodated. It means that the distribution of school funds to schools of different types should be very different in these two states. Again, in state A it is unlikely that more than 12 per cent, or about half the students who will attempt an academic course, will be able to finish the course and graduate. In state B, 44 per cent, or about two-thirds of all who enter, should be capable of completing the course, and should be provided for to the end of the course. That is, the proportion of ninth-year to twelfth-year students is likely always to be different in the two states; and again, a different apportionment of funds, this time within the school, is indicated. Further, in state A, not more than about 4 per cent of all the school children—about 1 in 6 of those who enter the academic high school—are likely to profit by taking algebra, as now taught. In state B, about 24 per cent or more than 1 in 3 of those who enter high school, may profit from the present algebra course. Therefore the subjects offered, or at least the number of pupils provided for, in each will need to be quite different in two such states.

Moreover, since in some of the southern states probably as many as 75 per cent of the children can not or will not enter academic high schools, the problem of providing other and perhaps new types of training for children from 14 to 18 years of age is most acute in this part of the country. There is here a fertile field for pioneer work in origina-

ting a curriculum which will fit their needs, for discovering what these children can and should be taught and what methods of presentation best reach them. What can best replace the academic curriculum for these children, to yield satisfaction in their own lives and enable them to become satisfactory citizens of a democracy? When educational authorities in the south see this as peculiarly *their* problem, and, with the increased federal aid which is coming, direct their efforts to solving it in their own way for their own region, rather than adopting the solutions of progressive western states where the proportions if not the conditions of the problem are quite different, we may expect new developments in secondary education which will command the attention of all.

Summary

1. Though Terman and others have indicated that mentality limits school achievement, few measurements are available to show just how fast or how far children of given mental equipment can progress in our schools.

2. The intelligence of the high school population in this country is limited to approximately the upper half of the whole range of American intelligence.

3 and 4. Intelligence is an important factor in determining the number of years a youth spends in school and college. The minimum intelligence usually necessary in order to enter high school is represented at age 14 by an Alpha score of 65; the minimum usually necessary to achieve high school graduation is represented at age 14 by a score of 85 points; the minimum for profit from present high school algebra is about 105.

5. Geographical differences in intelligence are enormous, the median for the lowest state being only half as great as that for the highest. In certain states more than half the population is below the level apparently necessary for academic high school work; in others, three-fourths of the population may be expected to enter high school. This has an important bearing on the distribution of school funds.

Bibliography

1. Terman, L. M.: "The Measurement of Intelligence." Houghton Mifflin, 1916.
2. Terman, L. M.: "The Intelligence of School Children." Houghton Mifflin, 1919.

3. Proctor, Wm. H.: Psychological Tests and Guidance of High School Pupils. *Journal of Educational Research Monographs*, No. 1, Public School Publishing Co.
4. *Annual Report*, Providence Public Schools, 1917–18.
5. Psychological Examining in the United States Army. *Memoirs of the National Academy of Sciences*, Vol. XV, 1921.
6. *Manual of Instruction for Use with the Army Alpha Intelligence Tests, Forms* V, VI, VII, VIII, IX *in Public Schools.* Kansas State Normal, Emporia, Kansas. 1922.
7. Doll, E. A.: The Average Mental Age of Adults. *Journal Applied Psychology*, December, 1919.
8. Madsen, I. N. and Sylvester, R. H.: High School Students' Intelligence Ratings according to the Army Alpha Test. *School and Society*, Oct. 4, 1919.
9. High School Scores on the Army Alpha Test. *Bulletin* of the Bureau of Tests and Measurements, Univ. of Mich., Feb. 7, 1921.
10. Terman, L. M.: Mental Growth and the IQ. *Journal of Educational Psychology*, September and October, 1921.
11. Goddard, H. H.: "Human Efficiency and Levels of Intelligence." Princeton Univ. Press, 1920.
12. Cobb, M. V. and Tape, H. A.: Note on a Method for Studying Causes of Increase in Alpha Scores. *School and Society*, June 24, 1922.
13. Thorndike, E. L. and others: "The Psychology of Algebra." Macmillan Co. (forthcoming).

ADDITIONAL DATA FROM CONSECUTIVE STANFORD-BINET TESTS

BIRD T. BALDWIN

AND

LORLE I. STECHER

State University of Iowa

This article presents supplementary data as a result of further tests by the Stanford Revision of the Binet Scale of 143 cases reported by the writers in January, 1922.[1] Just one year later 32 of the 36 cases who had received five previous examinations had a sixth; 40 of the 42 cases who had had four previous examinations had a fifth; 41 of the 51 cases who had had three previous examinations had received a fourth; 31 of the 56 cases who had had two examinations had received a third; 64 additional cases with two examinations were included.

Table I.—Coefficients of Correlation for IQ's. Boys and Girls

Examination number	1	2	3	4	5
2	+.850 ±.031				
3	+.738 ±.051	+.846 ±.031			
4	+.779 ±.044	+.802 ±.040	+.910 ±.019		
5	+.817 ±.037	+.815 ±.037	+.839 ±.033	+.918 ±.017	
6	+.812 ±.038	+.751 ±.049	+.796 ±.041	+.866 ±.028	+.944 ±.012

These new data confirm the findings of the previous study that for practical purposes the IQ remains sufficiently constant for a group as a whole, but that the individual records show fluctuations which are smoothed out in obtaining general averages. The amount of these fluctuations is evident in the tables of original data in the previous study, pp. 24–29, which have been brought up to date in mimeographed form and may be had on application to the writers.

[1] Baldwin, B. T. and Stecher, L. I.: The Mental Growth Curve of Normal and Superior Children Studied by Means of Consecutive Intelligence Examinations. *Univ. of Iowa Studies in Child Welfare*, 1922 (2), No. 1, pp. 61.

The inter-correlations with examinations, for those who have IQ's, (given in Table I), show the distribution of the individuals within this group on subsequent tests. The correlation between the fifth and sixth examination is the highest (+0.944), which probably means that the individuals have become thoroughly stabilized within the group.

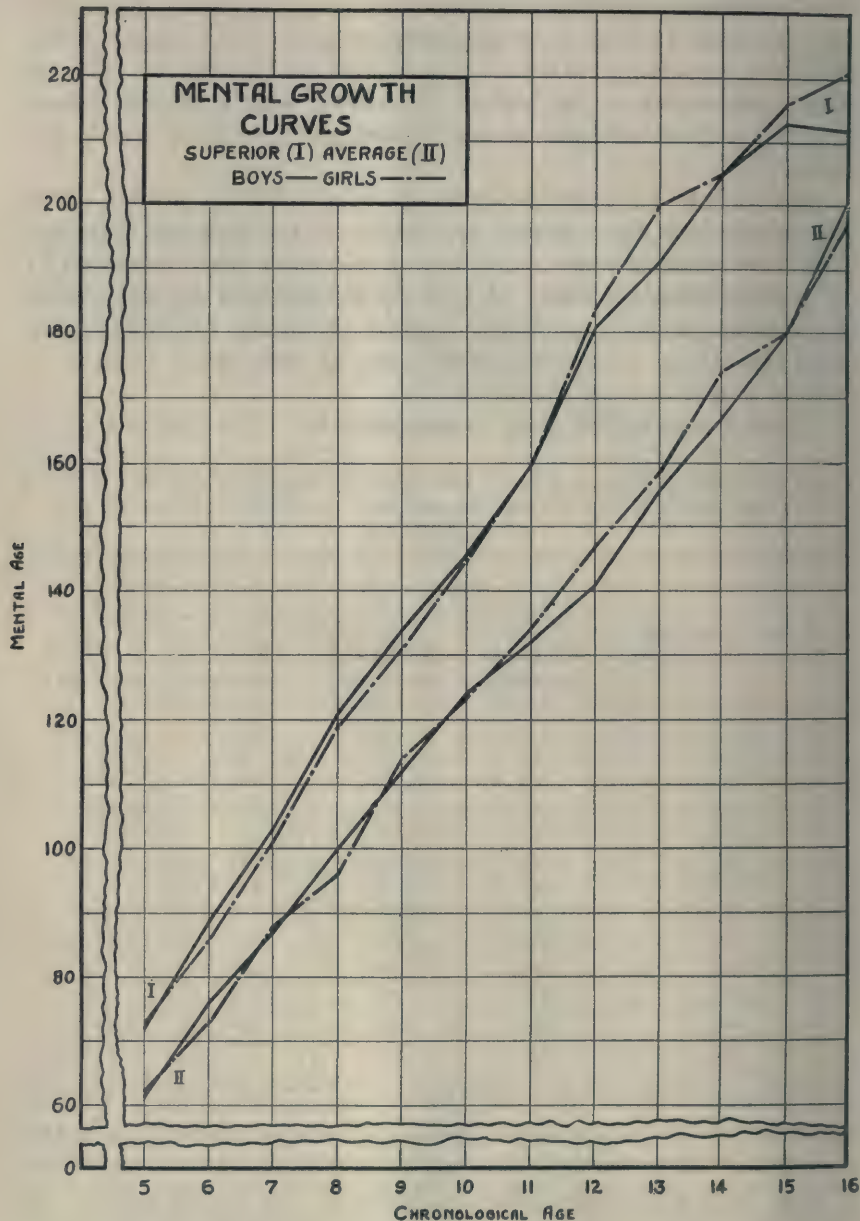
The writers have previously analyzed the sort of growth curve that results from the repeated application of the Stanford Revision. This curve represents one aspect of mental growth when measured by an existing tentative scale. Additional data permit the calculation (by the same method previously used) of the figures of Table II, the mean mental age in months for each sex at each age of children of superior and of average mental ability.

Chart 1 shows these data in graphic form. The curves have in

TABLE II.—MEAN MENTAL AGE IN MONTHS OF SUPERIOR AND AVERAGE BOYS AND GIRLS FOR SUCCESSIVE CHRONOLOGICAL AGES (BASED ON CONSECUTIVE EXAMINATIONS)

Chronological age	Boys		Girls	
	IQ 110+ (superior)	IQ 90–110 (average)	IQ 110+ (superior)	IQ 90–110 (average)
5	72	61	73	62
6	89	76	86	73
7	103	87	101	88
8	121	100	119	96
9	134	112	131	114
10	146	124	145	123
11	160	132	160	134
12	181	141	184	147
13	191	156	200	159
14	205	167	205	174
15	213	180	216	180
16	212	201	221	198

general the same appearance as those in the previous study with the exception of the curve for the average girls which lies much closer to the average boys curve than formerly, probably due to the addition of more average girls at this age. The average curves are approximately straight lines, which shows that these children are comparable to those


MENTAL GROWTH CURVES
SUPERIOR (I) AVERAGE (II)
BOYS —— GIRLS —·—
I
II
I
II
220
200
180
160
140
120
100
80
60
0
MENTAL AGE
5
6
7
8
9
10
11
12
13
14
15
16
CHRONOLOGICAL AGE

on whom the scale was standardized. In contrast with the straight-line average curves, the superior curves show fluctuations at the adolescent ages, indicative of the earlier mental development of superior children. Both the superior and the average girls of this group are in advance of the boys at the adolescent ages—12 to 14—when measured by this scale.[1] As previously pointed out, this adolescent spurt is analogous to the adolescent acceleration so frequently found in physical growth curves in height, weight, breathing capacity and other physical traits.

Unfortunately we have not, in the present state of development of the science, any measuring instrument that at all approximates the apparatus for measuring physical growth. The cheapest measuring stick is superior, both in equality of units and in extent, to our mental measurement scales. These poor mental tape lines wrinkle and stretch in places, and someone has cut off a little from both ends! The unit of measurement in mental growth scales is not an absolute unit such as the centimeter or the kilogram. The writers are in hearty agreement with the author[2] of a somewhat facetious review in regard to the desirability of discovering such an absolute unit of mental growth. An inch of growth in height is the same between 5 and 6 years or between 12 and 13 years. There is good reason to believe, however, that 2 months mental growth may mean a very different

[1] This conclusion has recently received some support from the evidence of Sullivan and Murdock (*Journal of Educational Psychology*, 1922, Vol. 13, 350–362).

[2] Sandiford, P.: *Journal of Educational Psychology*, 1922 (13) 378–379. The joint authors of this study, which the reviewer attributes mainly to one of them, take this opportunity to correct a few misapprehensions. (1) In view of the discussion above, there can be no objection to the plotting of mental age curves in regard to which the reviewer seems to have such a serious complex. (2) The reviewer comments on the fact that the authors believe the curves to be straight. That this is not the case is shown by the quotation (p. 12), "further analysis reveals, however, a very significant change in the trend with the approach of adolescence. This is especially marked in the curve for girls, etc." (3) The mental age curves and the IQ curves are, indeed, as the reviewer has aptly put it, "the same thing plotted in a different fashion." Although both are approximately straight lines, "there are fluctuations associated with physical development" (in the IQ curve) and "there is a significant change in the trend with the approach of adolescence" (in the mental age curve)—surely not, as the reviewer states, "diametrically opposite conclusions." (4) The authors presume that the reviewer failed to find one or two real errors which they now desire to point out. On page 12, beginning with line 25, one should read, "At 6 years -1 month, $+11$ months at the rate of 1.38 or $104 + (11 \times 1.38)$ or 119.18." Other proof-reading errors will be found on pages 12 and 17.

thing at these two periods. The amount of mental growth for 2 mental months at the earlier age may be only half that of 2 mental months at the later age. We do not know. We assume that the difficulty of the tests within the scale takes this into consideration and meets the differences fairly accurately. By the very fact of such construction, however, mental age scales tend to conceal any differences in the rate of mental growth that may exist. If any adolescent acceleration appears, it is all the more significant. Even the discovery of this hypothetical absolute unit of mental growth will not provide a scale for measuring mental growth, because mental growth like physical growth is a complex process involving development in a diversity of traits and functions. For example, physical growth is measured in inches, pounds, square inches, cubic inches, and a large number of other units for strength, temperature and metabolism measurements. It is possible to get some idea of the individual's development from a measurement of the height or the weight alone, but a complete growth curve is the result of composite measurements. That the writers have already pointed out this fact in the earlier study is shown by the following quotation (page 58), "Theoretically it would seem to be a better measure of mental growth to use a combination of point scales for specific mental traits, each scale to be sufficiently extended to measure whatever ability exists and the whole system to include a sufficient variety of traits to afford a general measure of the development of the individual."

A METHOD OF MEASURING FATIGUE OF THE EYES[1]

RALPH E. WAGER

Emory University, Ga.

Several factors are to be taken into account in connection with any method designed to measure fatigue of the eyes. There are involved: (1) The retina, (2) the refracting mechanism, and (3) the internal and external musculature.

Variations in retinal sensitivity may be followed in most cases by laboratory tests which need not be here described. It seems to be well established, however, that the purely nervous elements of the bodily mechanism are the last to suffer losses under adverse physical conditions. Not only so, but such tests do not lend themselves well to the measurement of fatigue because (1) wide variations occur within narrow time-limits, (2) central factors seem to be largely involved, and (3) certain physiological activities, such as the vaso-motor waves, or even the act of breathing, seem to influence the results.

The refracting mechanism may, in ordinary laboratory procedure at least, be regarded as a constant since changes take place in it but slowly and usually as a result of advancing years. Such changes are not within the range of the field covered by fatigue.

There remains, then, the musculature as the most significant factor. It is commonplace that muscular capacities are subject to marked changes due to physiological causes. This fact is obvious both through experience and experiment. The amount of change is in many cases measurable. One thinks, in this connection, of the work of Mosso and his epochal ergograph. Such recoverable losses in muscular capacity as are commonly experienced by all active muscles may be regarded as due to fatigue, and are accounted for as the result of the accumulation within the substance of the tissues of certain katabolic products. Recovery of power is due to either the transformation or elimination of these substances. Their presence, however, in sufficiently large amounts, brings about a lessened capacity to do work, largely, it appears, because of certain positively or negatively charged ions whose effect is to prevent the passage of an adequate stimulation into the muscle substance. In certain pathological cases it is likely

[1] A summary of a dissertation submitted to the faculty of the Graduate School of Arts, Literature and Science in the University of Chicago. The author acknowledges indebtedness to Dr. F. N. Freeman for suggestions and criticisms, and to many friends who assisted in the investigation.

that exhaustion of the energy-furnishing substances may also take place, but normally the inhibiting ions are developed in sufficient amount to manifest their effect before this condition is reached.

It is probable, therefore, that changes in ocular powers are due largely to changes in muscular capacities. The muscles involved are (1) the ciliaris, which controls the accommodative reactions, and (2) the external muscles which function in the acts of convergence and divergence. Both sets of muscles are brought into play when one shifts his field of regard from a point in a near plane to one more distant, or *vice versa*.

The method of measuring fatigue herein to be described is based upon the last made assumption. Its claim for merit lies chiefly in the fact that it is largely objective, and requires but little training or previous experience with the apparatus in order to be effectively used. It can be used with children who have learned to read.

The discussions to follow will be presented under the following topics:

1. Description of the apparatus.
2. Description of the tests and the manner of application.
3. Typical cases and discussion of results.
4. General conclusions.

Description of the Apparatus

The method, as has been suggested, is based upon the assumption that the most likely measure of ocular fatigue may be had by attempting to determine losses or gains in ocular muscular capacities resulting from rapid shifts of the field of regard so that both accommodation and convergence and divergence are necessitated. Such are involved when a change or shift in the field of regard is made from a near to a more distant plane, or the reverse, as has been suggested above.

In order to compel such shifts alternately from one plane to another, two hard-rubber discs, 6 and 8 inches in diameter respectively, were mounted on a steel rod 48 inches long. In front of each disc was a shield sufficiently large completely to cover it, and bearing in its upper margin an opening 1 inch square. Both shield and disc were movable along the rod whereon they might be secured at any point by means of a set-screw. The rod with its discs was supported upon suitable tripods and other accessory parts.

By means of parts which need not here be described in detail, the rod, with its attached discs, could be caused to rotate through a fraction

of a turn with each impulse applied to a foot-pedal. Each fractional rotation was through $\frac{1}{24}$th of a turn, or 15 degrees. These movements were controlled in extent, and made rapid and vibrationless, by means of suitable devices secured to, and operating against, a third disc attached to the distal end of the rod above mentioned. This latter was outside the point of attachment to the supporting tripod, and, by means of a lever and appropriate ratchets playing into two sets of toothed discs, controlled the partial rotations and at the same time, by means of pins set in its margin, made an electric circuit which served to register the instant of the movement of the discs.

Attached to the two discs first mentioned were paper forms of the same size, bearing on their margins printed words, either singly or in groups, so disposed as to bring each behind the opening on the shields at each partial rotation. These words were of different sized type for the near and distant plane so that the image formed upon the retina would be approximately the same size in each case.

The apparatus was placed upon tables so that, in the line of the two discs, a subject might be seated with his eyes about on a level with the openings in the shields. A suitable head-rest was provided. The near disc was placed at a distance of 8 to 10 inches from the eyes, and the more distant from 36 to 52 inches, depending in each case upon the accommodative capacities of the subject as determined by experiment. Shifts were then made in the planes as will be described below.

It is advisable at this point to make clear the manner in which the sequential fixations and accommodations were utilized as a basis of measuring changes in capacities.

It is evident that two essential factors are involved: (1) The *rapidity* with which the necessary muscular reactions are brought about, and (2) the *accuracy* with which they are accomplished. This latter involves the element of acuity, or correctness with which perception takes place. Each of these factors is taken into account in a manner next to be described.

The speed, or rapidity, with which the muscular adjustments were accomplished, was measured by (1) a voice-key, interposed between the subject and the near disc, and insulated against vibrations by being placed upon a sand-filled pedestal, and (2) the pegs on the rear disc (control) which served to make an electric circuit with each partial rotation. The voice-key was of the Roemer type, modified in a manner to make it automatic in its resetting, and exceedingly sensitive. The appearance of a word behind the opening of the shield was the

stimulus for its being read aloud as quickly as perception made it possible. The response of the voice-key to the sound waves made an electric circuit which recorded the instant of the subject's perception as evidenced by the spoken word, and the control disc made the contact which caused the registration of the instant of the appearance of the word, as above noted. Records were made on smoked paper in the form of a long belt stretched over two drums which were driven by a $\frac{1}{10}$ h.p. motor with suitable reducing and controlling elements to give a speed adapted to the purpose. A triple time marker was used. One element was operated by the make-circuit of the voice-key, another by the make-circuit on the control-disc, and the third by a Jacquet chronometer, placed in a circuit, for the purpose of giving a suitable time-record against which the other two might be projected and intervals determined. The time-record was made in fifths of seconds. The instant of appearance of the stimulus word being recorded, together with the subject's response thereto, it became possible to measure the period required for the fixation-accommodation act incident to it. This will be made clearer in connection with the explanation given below.

Acuity, was taken into account by noting the number of correct and incorrect responses. This will be referred to below.

Description of the Tests and Their Application

Two forms of tests were finally adopted as best suited to the purpose of the undertaking. These need to be described.

The first was called the 2-1 type. In it 24 words were used on the distant paper disc, and 12 on the near, each alternate space remaining blank. In applying the test, a word appeared in the opening of the rear shield, the near one being blank. Perception of the distant stimulus was followed by a partial rotation which brought a word behind each opening. The distant was then perceived and an immediate shift in the field of regard took place for the purpose of reacting to the near one. Another fractional rotation restored the original condition, following which the process was repeated until the complete series was run off. A portion of a record made in a test of this sort is shown in Fig. 1. The letter *A* indicates the registration of the instant of perception of the distant word; the partial rotation following is indicated at 1; the recognition of the new word in the distant plane is recorded at *B*, and the following in the near plane at *C*. The succeeding change in stimuli is marked by 2, which change restores the original condition.

Tabulations from a record of the above type are made in Table I in those parts indicated as I, I*a*, III and III*a*. Referring to I, the numbers in the column marked *a* is the time elapsing (in fifths of seconds) between the appearance of the stimulus word in the distant plane and its perception (1 to *B*); those in the column marked *b*, the

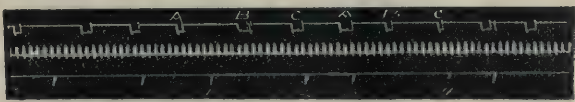


FIG. 1.

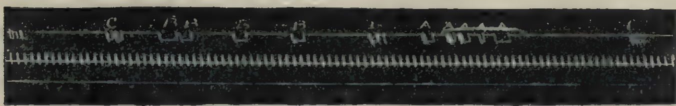


FIG. 2.

times for the fixations and accommodations necessary for the recognition of the word in the near plane (*B* to *C*); those in the column indicated by *c*, the times for the adjustments for the new word in the distant plane (*C* to *A*) the latter including also the reaction time of the operator (*C* to 2) which is given separately in column *d*. As will be described below, averages of each of these series of measures are required. In deriving that for column *c*, either the total of *d* is subtracted from *c*, or the average of *d* taken from that of *c* so that the effect of the operator's reaction time is eliminated. The averages of the columns (the function of which will be discussed below) are indicated as well as the average of the averages.

The second form of test used was called the 1-5-5 type. Its advantage (and difference) over the other lay chiefly in the fact that the adjustments in each plane had to be maintained for a longer time. On the rear disc a single word alternated with a series of five words. On the near disc, series of five words alternated with blanks. The test began with the recognition of a single word in the distant plane, the near opening being blank. A partial rotation followed, bringing behind each opening a group of five words. The more distant ones were read in as rapid succession as possible after which a shift in the field of regard took place to the near plane for the purpose of perceiving the words therein appearing, and which were likewise read as rapidly as possible. Another partial rotation brought a single word in the distant plane, thus restoring the original condition. Figure 2 is a

reproduction of a portion of a record made in connection with a test of this type. It is interpreted as follows: *C* marks the time of recognition of the single word in the distant plane. The succeeding fractional rotation is timed at I; the response to the five words in the distant plane are indicated at *B-B-B-B-B*, and to those in the near plane at *A-A-A-A-A*. The following partial rotation is timed at 2, and restores the original condition.

TABLE I

Subject Ph.
−2 lens
Distances: 9 and 36 inches
Stimulus lists:
I. *D* and *P* II. 5*y* and 5*c* III. *E* and *L*
Ia. I*a* and *Q* II*a*. 5*x* and 5*a* III*a*. *A* and *R*

	(*a*)	(*b*)	(*c*)	(*d*)		(*a*)	(*b*)	(*c*)	(*d*)	(*e*)
I	4.5	9	10.5	3	II.	8	25	12	23	8
	5	7	24	16		6.5	20	10.5	15	6
	6	10	9	3		8.5	34	12	14	11
	7	16	10	5		10	14	13	13	8.5
	8.5	7	11	2		6.5	19	11.5	13	6.5
	5	7.5	16	2.5		8	24	13.5	14.5	16.5
	5.5	8	18	2		7.5	10	13.5	15	21
	5.5	11.5	11	4		15	35	14.5	15	18
	7	9	9	3		9.5	43	38	71	14
	14	10	9	2.5		21.5	31.5	18	15	23
	5.5	9	11	2.5		15	10	14	9	17
	6	12	9.5	3		19	37	31	47	10
	79.5	116	148	48.5		135	302.5	201.5	264.5	149.5
	6.6	9.7	12.3	4		11.2	25.2	16.8	22	12.5
Average, 8.2 Errors, 3					Average, 17.5 Errors, 21					
Coefficient, .65					Coefficient, .24					
I*a*.	26	24	23.5	8	II*a*.	9	19	8	26	5.5
	7	9	18	6.5		9	44	13	22	11
	7	6.5	17	9		12	72	23	38	11
	8	8	16.5	2		8	48	12	28	7
	9	9	11	3		17	25	12.5	14	11
	12	9	32.5	20		18	40	12	17.5	14
	13	9	14	5		12	40	11	16	23
	6	7	18.5	9		17.5	46	13	14	12
	7	6.5	27	7		14	30	11	12	14
	21	12	11	3		9	23	8.5	8	10
	19	14.5	44	13.5		14	33	8	13	8
	8.5	10.5	8.5	2						
	143.5	125	251.5	88		148.5	436	143.5	227.5	133.5
	11.9	10.4	20.9	7.3		12.4	36.3	11.9	18.9	11.1
Average, 11.9 Errors, 5					Average, 18.1 Errors, 28					
Coefficient, .36					Coefficient, .21					
III.	5.5	9	13.5	2	III*a*.	8	8.5	10	3	
	6.5	10.5	26	15		7.5	11	20	3	
	6	20	19.5	7.5		10	15	27	13	
	15	12.5	27	19.5		7.5	21	12	3.5	
	7	8	11	3		7	11	11.5	4.5	
	6	12	14.5	6		8	9	31	4	
	6.5	9	11	2.5		11	8	13	4	
	6	10.5	13.5	2.5		6	8	16	2.5	
	10	11	11	2		16	10	11	3	
	8	10	20	6.5		13.5	12	22	14	
	7	14	23	7		7	7	9.5	3	
						10	8	12	2.5	
	83.5	126.5	190	73.5		121.5	128.5	195	60	
	7.6	11.5	17.3	6.7		10.1	10.7	16.2	5	
Average, 9.9 Errors, 8					Average, 10.6 Errors, 3					
Coefficient, .38					Coefficient, .43					

Average loss, 18.5 per cent.

Tabulations made from the type of test above described are shown in Table I, Parts II and II*a*. Referring to the column marked (*a*),

it is to be understood as made up of the times for the perceptions of the first of the five words in the distant plane (1 to *B*); column *b*, the times for the similar reactions to the remaining four words of the series (*B-B-B-B*); *c*, the times for the adjustments involved in shifting from the distant to the near plane (*B* to *A*); *d*, the times for the reading of the remaining four words of the near series (*A-A-A-A*); *e*, the times for the perception of the single word in the distant plane following a change in stimuli (2 to *C*). Averages of the columns are also given as is also an average of the averages. It is to be noted that columns *b* and *d* are for the intervals involved in reacting to four successive stimuli so that as a result the final average of this type of test is not comparable with that of the 2-1. It would be possible to measure each reaction separately, but the manner in which the measures were utilized made this unnecessary.

Another important factor in the test remains to be noted. In order to place the muscles of accommodation upon their near-limit, and thus make fatigue effects the more quickly and certainly manifest, the subject wore, during the test, minus lenses of a refracting power sufficiently great just to enable adjustments in the two planes. A feeling of effort accompanies such a condition. The lenses used were determined by careful experimentation with each subject. This was usually done in connection with the preliminary trials with the apparatus.

It was essential that the outcomes of the tests be reducible to single number which might serve as a measure of the ocular powers. Thus only might valid comparisons be made. Note has been made of the fact that the tests involve both speed (rapidity in fixations and accommodations) and acuity (correctness of perceptions). Measures of each were combined into a single number in a manner next to be explained.

Since both factors are clearly involved in effective ocular activities, the final measure must reflect variations in either one. This necessity was met by making the percentage of correct perceptions the numerator of a fraction whose denominator was the average fixation-accommodation time (the average of the averages) expressed in hundredths of seconds. This ratio is obviously influenced by changes in either factor. It may be that one is more significant than the other, but as yet only speculation is possible on the point. The fraction, reduced to the form of a decimal, was called the fixation-accommodation coefficient. It seems, from the evidence collected, to be a fair means of representing numerically the ocular capacities existing at the time of

the test. In Table I such coefficients are indicated for each type of test used in an actual series of such measures.

It became apparent as the result of repeated applications of the tests to individual subjects, that each tends to exhibit a characteristic fixation-accommodation time (within limits). Since this is a factor in determining the value of the coefficient, it is clear that comparisons for the purpose of determining relative gains or losses are valid only when made between coefficients belonging to the same individual. Initial, or native, capacities may be represented with at least a certain degree of validity by means of the coefficient. Because of such differences valid comparisons for the purpose of noting changes can be made only between individual scores. Furthermore, the fact that the average of the 1-5-5 type was based upon the use of a single measure for the reading of the four words in both near and distant planes gave to the coefficient derived therefrom a value not at all directly comparable with that from the 2-1 type. Comparisons can be made only between tests of the same type. The manner in which this was done is explained below.

Application of the Tests

After the tests were developed it became advisable to accumulate evidence as to their reliability. This was done by applying them to subjects before and after ocular fatiguing experiences and noting the manner in which the results compared with the expectations. Owing to limited space, only a brief account can here be given concerning either the detailed tests or the applications to individual subjects. The following account, however, may serve to indicate the nature of the work done, and the general conclusions may be accepted as based upon the evidence. The number of cases as yet investigated is too limited to yield highly significant conclusions. The investigation was concerned primarily with the evolution of a method, and not so directly with the development of a body of facts based upon its application. However, certain facts seem obvious from the cases studied, and they are stated in the concluding paragraphs. We shall next describe the manner in which the tests were applied.

At some time following the preliminary experience with the apparatus, and the determination of the lenses to be worn, a battery of tests was given to the subject as follows: A 2-1, a 1-5-5, and a 2-1 type test in the order named. If it were desired that ocular fatigue be induced, for the purpose of attempting to measure it, the subject was

immediately placed in an illumination so low as to place a severe burden upon the eyes, and in it he read for a stated period. The degree of illumination was measured by an illuminometer. In other cases, normal conditions of illumination prevailed, according to the purpose of the investigation. In either event, the reading was followed by the administration of another battery similar to the first and given under identical conditions of illumination, etc. For each test the coefficient was then determined. Those preceding the reading were called antecedent tests, and those following, subsequent. The coefficients of the antecedent series were then compared with the corresponding member of the subsequent and the percentage of gain or loss, as evidenced by the coefficients, was determined for each pair. The average of the three results (summed algebraically) was used as the most likely measure of the losses or gains. Reference to Table I will clarify the procedure.

As an illustration of the manner in which the tests were used, and also for the purpose of presenting some of the evidence on which the conclusions are based, the following typical cases will be cited.

Sujbect Ob. Positive Accommodation, 3.5 diopters

Test 1.	Read 1 hour 15 minutes in .8 f.c. illumination.	Loss 10.8 per cent
Test 2.	Read 1 hour 55 minutes in .5 f.c. illumination.	Loss 11.8 per cent
Test 3.	Read 1 hour 10 minutes in 28 f.c. illumination.	Loss 19.9 per cent

Test 3 was made at 10:00 A. M. The subject reported that his eyes "felt tired" due to late reading the night preceding by way of preparation for an examination. This may explain the large loss following reading in a favorable illumination. Evidently the fatigue of the excessive use of the eyes carried over until the following day. Another case, almost identical, but with even greater losses, adds to the validity of the explanation used.

Subject Ph. Positive Accommodation, 2 diopters

Test 1.	Read 1 hour 10 minutes 1 f.c.	illumination.	Loss 18.5 per cent
Test 2.	Read 1 hour 40 minutes .1 f.c.	illumination.	Loss 18.6 per cent
Test 3.	Read 1 hour 33 to 76 f.c.	illumination.	Gain 13.7 per cent

The gain in Test 3 is to be noted. It was made early in the morning before any study had been done. It is possible that a sufficiently pronounced "warming up" was not given before the antecedent series. This subject in these and other tests exhibited a marked susceptibility to fatiguing conditions. Note should be made also that both he, and several others, reported a peculiar rhythmic appearance

and disappearance of the word which served as stimulus following fatiguing experiences. The effort to see clearly was frequently of little avail. The perception came "on the fly." Others described the sensation as like that which would result if the words were on a spring bobbing slowly forward and backward. This was due, undoubtedly, to the rhythmic contraction and relaxation of the ciliari.

Subject Br. Positive Accommodation, 2.25 diopters

Test 1.	Read 2 hours 15 minutes	1 f.c. illumination.	Loss 18 per cent
Test 2.	Read 1 hour 40 minutes	.5 f.c. illumination.	Loss 12 per cent

This subject had a very high-pitched and feeble voice. It was possible for her, however, after a little training, to speak against the voice-key in such a manner as to make good records.

Subject O. Positive Accommodation, 3 diopters

Test 1.	Read 1 hour 10 minutes	1 f.c. illumination.	Loss 23 per cent
Test 2.	Read 1 hour 10 minutes	68 f.c. illumination.	Loss 7.5 per cent

This subject lost even in good illumination. He has excellent vision, but is susceptible to fatigue. He reported trouble with his eyes whenever he has had to read under poor illumination.

Subject Her. Positive Accommodation, 3 diopters

Test 1.	Read 55 minutes	.5 f.c. illumination.	Gain 61 per cent
Test 2.	Read 1 hour	.3 f.c. illumination.	Gain 14.6 per cent
Test 3.	Read 2 hours 40 minutes	.2 f.c. illumination.	Gain 10.8 per cent

The fact that a gain occurred in each test is a noteworthy fact. It is to be observed, however, that with an increase in the reading period and a decrease in the illumination, the amount of gain is correspondingly lessened. There is evidence that the gain in power is genuine. The case next to be presented is similar.

Subject Kl. Positive Accommodation, 4.5 diopters

Test 1.	Read 1 hour 15 minutes	.2 f.c. illumination.	No change
Test 2.	Read 2 hours	1 to 1.5 f.c. illumination.	Gain, 5 per cent

This subject, like the preceding, manifested a marked resistance to fatiguing conditions. A test other than the two here reported was made by him, and yielded similar results.

It is appropriate to add at this point certain further comments concerning the general nature of the results.

It is apparent that there is a wide variation in the susceptibility of subjects to fatiguing conditions. Some, of which the first four

cases presented may be chosen as typical, tend to show a marked falling off after relatively short reading under low illuminations. Others, such as the last two cases, are highly resistant. In three of our cases well marked gains occurred after periods of reading under conditions which one would expect to reduce greatly the ocular powers. To the former type we have given the term "minus" since the losses are immediate and marked; to the latter, the term "plus" since there is frequently a gain following the fatiguing conditions. These two represent the extremes of a varying series. That there is an actual increase in ocular capacities is evidenced not only by the tests, but also from subjective evidence. One subject in particular reported a feeling of exhilaration and ability to distinguish words which before the reading could not be seen clearly. It may rightly be argued that this sensation is itself good evidence for fatigue. That the eyes were not impaired in their functions is, however, the important fact.

That these differences are not ascribable to initially higher powers on the part of the one type is evidenced by a consideration of the coefficients of the antecedent series of the two types. Such a comparison is shown in the following figures:

	Average Coefficients Antecedent 2-1 tests		Average Coefficients Antecedent 1-5-5 tests
Plus types	.44	.62	.46
Minus types	.57	.60	.54

From this it appears that there is no striking initially superior capacities belonging to the more resistant individual.

These extremes differ with respect to one characteristic which may in part account for the results they yield. If the average deviation of each of the series of measures (*i.e.*, the columns) be determined, and that of the antecedent then compared with the corresponding subsequent element, it is found that the minus type tends to exhibit a more pronounced increase in variability. The coordinations are obviously less perfectly controlled. Such a comparison is presented below.

Plus types

22 paired measures, of which 13 show decreases, 9 increases

Minus types

33 paired measures, of which 14 show decreases, 19 increases in the average deviations.

Comparisons were also made as to the effect of fatigue upon the coordinations involved in the adjustments for the shifts in the two directions—toward and away from the subject. Here again it appears

that the minus type exhibits the more marked increases in average deviations, especially in the shift from the near to the distant plane. This difference may indicate nervous excitability, or actual muscular incapacity. It is conceivable too, that central factors are concerned. Possibly all are.

Without attempting to present the detailed evidence upon which they are founded, a summary of the results of the application of the tests, as well as the implications from the conditions under which they were given, may be brought together in the following

General Conclusions

1. Ocular fatigue may be measured by the speed of shifts in fixation from a near to a more distant plane, and by the accuracy of the perceptions, the process being accompanied by the wearing of minus lenses fittingly chosen. The fixation-accommodation coefficient derived from the records thus made is evidently a fair measure of ocular capacities.

2. The validity of the method is evidenced in that the results obtained by its use correspond in the main with the expectations.

3. From the cases investigated it appears that ocular powers range from those (1) in whom a high resistance to fatiguing conditions prevails to those (2) in whom a ready susceptibility exists. These extreme types we have called the "plus" and "minus."

4. Individuals differ greatly with respect to their average fixation-accommodation times. A characteristic average time (within limits) appears to exist.

5. Fatigue tends to increase the variableness of the muscular adjustments in the less resistant types.